1984

Admissibility in choosing between experiments with applications

Reda Ibrahim Mazloum
Iowa State University

8423726

Mazloum, Reda Ibrahim

ADMISSIBILITY IN CHOOSING BETWEEN EXERIMENTS WITH APPLICATIONS

Iowa State University

PH.D. 1984

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Admissibility in choosing between

experiments with applications

by

Reda Ibrahim Mazloum

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1984

TABLE OF CONTENTS

1. INTRODUCTION

Suppose that X is a random variable whose density function is indexed by a parameter θ where $\theta \in \Theta$. For estimating some parametric function, say $\tau(\theta)$, the usual interest is to obtain an admissible estimator where an estimator is defined to be admissible as follows:

Definition 1.1:

An estimator δ is said to be admissible if there does not exist any other estimator δ^o such that $r(\delta^o;\theta) \leq r(\delta;\theta)$ for all $\theta \in \Theta$ with strict inequality for at least one θ where $r(\delta;\theta)$ is the risk function of δ when θ is the true state of nature.

The Bayesian procedure that uses a prior distribution to obtain a unique Bayes decision rule has been widely used as a tool to obtain admissible decision rules. For example, in the simplest case when Θ is finite and the prior distribution puts positive mass on every $\theta \in \Theta$ and when the loss function is such that the resulting Bayes decision rule is unique, then this rule is admissible. However, the class of unique Bayes decision rules doesn't form a complete class. In other words, there are some cases where a decision rule is admissible, but there does not exist a prior distribution against which this admissible rule is unique Bayes. To see that we consider the following example:

Example 1.1:

Let X be binomial (n,θ) where $\theta \in [0,1]$. For estimating θ with squared error loss, the estimator $\delta(X) = \frac{X}{n}$ is admissible

(Lehmann (1983)). However, the only way for $\delta(X)$ to be Bayes against some prior is that the prior assigns probability 1 to the set $\{0,1\}$ since if the prior assigns positive mass to any parameter point other than or besides $\theta = 0,1$, then the resulting Bayes estimator will not agree with $\delta(X)$ on $X = 0,n$. Therefore, $\delta(X)$ cannot be unique Bayes against any single prior.

On the other hand, when the Bayesian procedure uses a prior that yields a class of Bayes rules rather than a unique one, this class usually contains both admissible as well as inadmissible decision rules. To illustrate that, we consider the following example:

Example 1.2:

Let X be binomial $(3,\theta)$ where $\theta \in \{0, 0.4, 0.5, 1\}$. For estimating θ with squared error loss, consider the prior distribution, say λ^1, which puts mass 1/2 on 0 and 1/2 on 1. It is easy to see that any decision rule δ with $\delta(0) = 0$ and $\delta(3) = 1$ is Bayes against λ^1. In particular, the decision rules δ' and δ'' where $\delta'(0) = 0$, $\delta'(1) = 1/3$, $\delta'(2) = 2/3$ and $\delta'(3) = 1$ and $\delta''(0) = 0$, $\delta''(1) = 0.446$, $\delta''(2) = 0.457$ and $\delta''(3) = 1$ are both Bayes against λ^1. However, δ'' as we will soon see, is admissible while it is easy to show that δ' is inadmissible since it is dominated by δ''.

Now, suppose in example 1.2 we define a second prior, say λ^2, which puts mass 1/2 on $\theta = 0.4$ and 1/2 on $\theta = 0.5$ and we compute the Bayes estimates for those x's for which the Bayes estimate under

λ^1 is not defined, i.e., for $X = 1,2$ then we will get $\delta(1) = 0.446$ and $\delta(2) = 0.457$. Combining these two estimates with those defined under λ^1, it seems plausible that the resulting estimator δ'' is admissible. That is, by considering a second prior we could have extracted an admissible rule from the class of Bayes rules against λ^1. In fact, this idea of using stepwisely a set of mutually orthogonal priors (i.e., with mutually exclusive supports) to get a decision rule, called a "stepwise Bayes decision rule" is well-known and has been discussed in the literature (see Hsuan (1979), Meeden and Ghosh (1981), and Brown (1981)). For example, Meeden and Ghosh (1981) have given a minimal complete class theorem in the case when both the parameter space and the sample space are finite and the loss function is such that the prior risk under any prior distribution is uniquely minimized by a member of the decision space. This theorem says that "a decision rule is admissible if and only if it is unique stepwise Bayes against a set of mutually orthogonal prior distributions." According to this theorem, any admissible decision rule is unique stepwise Bayes. For instance, the admissible estimator $\delta(X) = \frac{X}{n}$ (which is shown in example 1.1 to be not unique Bayes against any single prior distribution) is unique stepwise Bayes. To see that, let λ^1 be a prior distribution that puts its mass on the set $\{0,1\}$ then it is easy to see that the resulting Bayes rule, say $\delta^o(X)$, is such that $\delta^o(0) = 0$ and $\delta^o(n) = 1$. Now, define the prior distribution λ^2 such that $\lambda^2(\theta) \propto [\theta(1-\theta)]^{-1}$ for $\theta \in (0,1)$ and 0 otherwise. (Note λ^1 and λ^2 are orthogonal). The resulting Bayes estimator is $\delta^{o}(X) = \frac{X}{n}$ for $X \in \{1,2,\ldots,n-1\}$. Hence, the unique stepwise Bayes rule against λ^{1} and λ^{2} is $\delta^{o}(X) = \frac{X}{n}$ which is $\delta(X)$. On the other hand, to obtain an admissible decision rule it is enough to obtain a unique stepwise Bayes decision rule. For example, we have seen at the beginning of this paragraph that $\delta''(X)$ given in example 1.2 is unique stepwise Bayes against λ^{1} and λ^{2}. Hence, by the above theorem $\delta''(X)$ is admissible.

A philosophical interpretation of using more than one prior may not be palpable. However, Hsuan (1979) has given the following result regarding the use of more than one prior: Consider a decision problem with a finite parameter space and a strictly convex loss function. Let $\{\lambda_n\}$ be a sequence of prior distributions where each prior is supported on the entire parameter space. Then, there exists a set of finitely many priors such that the limiting Bayes rule (assuming the existence) against $\{\lambda_n\}$ is identical with the unique stepwise Bayes rule obtained by using the finite set of priors stepwisely. In other words, this result says that a set of finitely many orthogonal priors is equivalent to a set of countably many priors, each of which is supported on the entire parameter space.

As shown by Meeden and Ghosh (1983), the stepwise Bayes procedure can also be used to study admissibility in the case of choosing among several experiments. For example, suppose that a statistician is interested in making some decision about some parametric function, say $\tau(\theta)$ where $\theta \in \Theta$. Before making his decision, he has the chance to observe, possibly at random, one of two possible experiments X_1 or X_2 where the probability function for each experiment is indexed by θ. Suppose that for $i = 1,2$, δ_i is a possible decision function to be used in connection with X_i and γ_i is the probability of observing X_i where $\gamma_i \geq 0$ and $\sum_{i=1}^{2} \gamma_i = 1$. Now, the question is which pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ where $\underset{\sim}{\gamma} = (\gamma_1,\gamma_2)$ and $\underset{\sim}{\delta} = (\delta_1,\delta_2)$ should be chosen by the statistician? Intuitively, the statistician would like to choose a pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ that cannot be dominated by any other pair. This suggests the following natural definition of admissibility of $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$:

<u>Definition 1.2</u>:

A pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is said to be admissible if there does not exist any other pair $(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^o)$ such that $r(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^0;\theta) \leq r(\underset{\sim}{\gamma},\underset{\sim}{\delta};\theta)$ for all $\theta \in \Theta$ with strict inequality for some θ where $r(\underset{\sim}{\gamma},\underset{\sim}{\delta};\theta) = \sum_{i=1}^{2} \gamma_i r_i(\delta_i;\theta)$.

When faced with such a problem, it is well-known that a Bayesian would choose a prior and compute the corresponding Bayes decision rules δ_1 and δ_2 for the experiments X_1 and X_2 and then choose the experiment with the smaller Bayes risk. He would only consider randomly choosing between the two experiments when the two Bayes risks are equal.

If he follows the method outlined above and it was the case that δ_1 and δ_2 are unique, then it is easy to show that his resulting pair will always be admissible. However, the class of all pairs obtained in this way does not form a complete class. On the other hand, if at least one of δ_1 and δ_2 is not unique then the Bayesian will get a class of pairs rather than a unique one. This class usually contains both admissible and inadmissible pairs.

For the above reasons, Meeden and Ghosh (1983), using the idea of "stepwise Bayes rules," have provided a technique to characterize the admissible pairs for this problem. This technique, when both the parameter space and the sample space are finite and the loss function is such that the prior risk is uniquely minimized by a member of the decision space, is summarized in a minimal complete class theorem. According to this theorem, the statistician can obtain an admissible pair for the above set up by choosing a sequence of mutually orthogonal priors $\lambda^1,\ldots,\lambda^m$ that yields unique stepwise Bayes rules δ_1 and δ_2 for X_1 and X_2 and computing the Bayes risks under λ^1. If the Bayes risks are not equal, he chooses the experiment with the smaller Bayes risk. If they are equal, he computes the Bayes risks under λ^2. If the Bayes risks under λ^2 are not equal then he chooses the experiment with the smaller Bayes risk. Otherwise, he computes the Bayes risks under λ^3 and so on until there exists a j^* where $1 \leq j^* \leq m$ such that one of the two Bayes risks under λ^{j^*} is smaller than the other. If this is the case, then the statistician chooses the experiment with the smaller Bayes risk. If not, then it must be the case that the Bayes risks are equal under all the priors. In this case, the statistician can choose between the two experiments in any way he wants.

Conversely, every admissible pair can be obtained using this theorem.

So far we have presented, for the purpose of simplicity, everything in the case of two experiments with the same parameter space. However, as shown by Meeden and Ghosh the extension to the case of k experiments with the same (or different) parameter space(s) is true.

Looking at the above set up for choosing between two experiments, we see that it was assumed that the class of all discrete probability measures, say $\Gamma = \{\underset{\sim}{\gamma}\}$ where $\underset{\sim}{\gamma} = (\gamma_1, \gamma_2)$ defined on $\{1,2\}$ was available to the statistician to choose from and, therefore, any admissibility results will be relative to Γ. But this is not always the case, i.e., in some cases the statistician will, for some reason, find himself restricted to choose from a subclass of Γ, say Γ^*. For example, suppose that observing X_i costs c_i units where $c_1 < c < c_2$ and that the statistician is willing to use any $\underset{\sim}{\gamma} = (\gamma_1, \gamma_2)$ as long as the expected cost is not more than c. In this case, the class of discrete probability measures available is no longer Γ, it is a restricted subclass of Γ namely, $\Gamma_c = \{\underset{\sim}{\gamma}\colon \sum_{i=1}^{2} \gamma_i c_i \leq c\}$. Therefore, our interest is to characterize the admissible pairs $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ relative to Γ^* where Γ^* is any subclass of Γ, for example, Γ^* could be Γ_c or even Γ itself.

In Section 2.1.1, adding this restriction to the set up given by Meeden and Ghosh (1983), a theorem that characterizes the admissible pairs relative to any arbitrary Γ^* is given in the general case of k experiments. In the case when $\Gamma^* = \Gamma$, the results of Section 2.1.1

yield the earlier results.

Now, suppose that the statistician is interested in making his decision based on selecting a subset of experiments rather than a single one. In this case, we note that the selection of that subset could be made nonsequentially or sequentially. For this reason, we devote Sections 2.1.2 and 2.1.3 to show how the results of Section 2.1.1 can be used to give a characterization of the class of admissible pairs in the two cases of nonsequential and sequential selection respectively.

So far, everything has been presented in the case when the parameter space is finite. However, obtaining an admissible estimator (or pair) using the stepwise Bayes procedure is not contingent on the finiteness of the parameter space. In other words, if the parameter space is not finite and it is easy to define a sequence of mutually orthogonal priors on it, then an admissible estimator (or pair) can be obtained using the method outlined before. However, in some cases it is not easy to define a set of mutually orthogonal priors on an infinite parameter space. For example, as we will soon see, in decision problems in finite population sampling the parameter space is usually taken to be $\mathbb{R}^N$, the N dimensional Euclidean space, and it is not easy to define a set of mutually orthogonal priors on such a parameter space. For this reason, Meeden and Ghosh (1982) introduced a concept called "finite admissibility." The basic idea of this concept is to have admissibility on every finite subset of the parameter space. More precisely, a decision rule δ (or a pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$)

is said to be finitely admissible if for any parameter point θ_o there exists a finite parameter subset Θ_o containing θ_o such that when Θ_o is taken as a restricted parameter space, δ (or $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$) is admissible. Moreover, they have shown that every finitely admissible decision rule (or pair) is admissible.

A well-known area of applications of admissibility is finite population sampling: Suppose that in a population consisting of N units the interest is to estimate, with squared error loss, some parametric function, say $\tau(y)$, where $y = (y_1,\ldots,y_N)$ is a vector of N population values of some characteristic of interest and is assumed to belong to $\mathbb{R}^N$, the N dimensional Euclidean space. A design p is a discrete probability measure defined on the set of all possible samples from this population. If δ is an estimator of $\tau(y)$ then it is interesting to ask if δ is admissible when a design p is used where p belongs to some class of designs say $\mathcal{P}$. A more interesting question is whether or not there exists another pair (δ',p') with $p' \in \mathcal{P}$ such that (δ',p') dominates (δ,p). If there does not exist such a pair then (δ,p) is said to be uniformly admissible relative to $\mathcal{P}$.

Now, by considering the samples to be the experiments available to the statistician, we see that the above questions are of the type discussed in this chapter. Since the parameter space is $\mathbb{R}^N$, the concept of finite admissibility is so useful in proving admissibility and uniform admissibility in finite population sampling. In fact, using this idea, Meeden and Ghosh (1982 and 1983), Ghosh and Meeden (1982), and Vardeman and Meeden (1983a, 1983b and 1984) have given various admissibility results in finite population sampling.

If the interest is to study uniform admissibility relative to the class of designs of fixed sample size n then the theorem of choosing between experiments given by Meeden and Ghosh (1983) can be used by considering the set of all samples of size n along with the class of all designs defined on that set. However, if the interest is to study uniform admissibility relative to some other class, say the class of designs of expected sample size n, then there is not any set of samples so that when used in connection with the class of all designs gives the class of designs of expected sample size n. For this reason, the theorem given in Section 2.1.1 is used in Chapter 3 to give some uniform admissibility results relative to some classes of designs.

For estimating the population total, the uniform admissibility relative to the class of designs of expected sample size less than or equal to n of some different strategies is demonstrated in Section 3.1.2. From those results, the uniform admissibility of those strategies relative to the class of designs of fixed sample size n follows easily.

In Section 3.1.3, following the line of argument given in Ghosh and Meeden (1982), an admissible estimator U^* of a parametric function U_p is constructed where this function U_p is the finite population sampling counterpart of a U-statistic. This class of functions contains as special cases the population mean, the population variance and many others. It turns out that U^* is just an appropriate multiple of the U-statistic corresponding to U_p defined on the sample. In the special case when U_p is the population total, U^* turns out to be the classical estimator which was first proved to be admissible in Joshi (1965). Also, in the special case when U_p is the population variance, the admissible estimator U^* obtained here was first constructed in Ghosh and Meeden (1982).

Godambe (1969) has given a uniform admissibility result relative to the class of designs of expected sample size n when estimating the population total. This result can be alternatively proved using the ideas of Chapter 2.

As shown by Meeden, Ghosh and Vardeman (1984) another application of utilizing the stepwise Bayes procedure in studying admissibility questions is in nonparametric problems. For instance, suppose $X_1,\ldots,X_n$ is a random sample from an unknown distribution F which is assumed to belong to Θ, some nonparametric family of distribution functions. For estimating, with squared error loss, $\tau(F) = \int\psi(t)dF(t)$ where ψ is some specific function. Meeden, Ghosh, and Vardeman (1984) have shown that admissible estimators for $\tau(F)$ can be obtained by considering only the subfamily of Θ consisting of all distribution functions which concentrate all their mass on a set of r distinct real numbers $\alpha_1,\ldots,\alpha_r$. If for every choice of $\alpha_1,\ldots,\alpha_r$, an estimator δ is shown to be unique stepwise Bayes for τ, then it is admissible for this simpler problem and hence it is admissible for the nonparametric problem as well. Moreover, they have shown that there is a natural duality between admissible estimators in the nonparametric problem and admissible estimators in the finite population sampling problem.

Using this idea of reducing the nonparametric problem to a simpler one, we show in Section 3.2 that uniform admissibility results can be obtained as well. For estimating $\tau(F)$, a uniformly admissible pair $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ relative to the class of designs of expected sample size less than or equal to n is obtained where $\underset{\sim}{\delta}$ is the unique stepwise Bayes estimator against some sequence of mutually orthogonal priors and $\underset{\sim}{\gamma}$ is the design that chooses the random sample of size n with probability one. In Section 3.3, using the duality, obtained by Meeden, Ghosh, and Vardeman (1984), between admissible estimators in the nonparametric problem and admissible estimators in the sampling problem we show that there is a corresponding duality between the Bayes risks in the two problems. In the case of having a prior distribution that yields a unique Bayes rule for the nonparametric problem, this Bayes risk duality leads to a uniform admissibility duality between the two problems and consequently, uniform admissibility in one problem can be obtained by knowing it in the other. In particular, if δ is unique Bayes for the nonparametric problem based on a random sample of size n then the corresponding estimator in the sampling problem along with any design of fixed sample size n is uniformly admissible relative to the class of designs of fixed sample size n, i.e., uniform admissibility relative to the class of designs of fixed sample size n for the sampling problem can be obtained by only knowing a unique Bayes estimator based on a random sample of size n in the nonparametric problem.

2. ADMISSIBILITY IN CHOOSING BETWEEN EXPERIMENTS

Suppose that a statistician is faced with a decision problem about an unknown parameter $\theta \in \Theta$ where Θ is finite. Before making his decision he can observe, possibly at random, one of k possible experiments $E_1,\ldots,E_k$ with corresponding random variables $X_1,\ldots,X_k$ where the probability function for each experiment is indexed by θ. The problem for the statistician is how to choose between these experiments. The answer to this problem is given by Meeden and Ghosh (1983). This answer implicitly says that essentially the statistician need not randomize in his choice. In Section 2.1.1, we will be looking at this problem in the case when the statistician is no longer able to choose freely among those experiments. In other words, in some cases the statistician will find himself, for some reason like cost or time limitation, restricted in his choice. Considering this restriction we find that the statistician sometimes needs to randomize in a specific way in choosing among those experiments since choosing with probability one a specific experiment might not be available to him.

In Section 2.1.2, we will be considering this problem when the interest is to choose nonsequentially n out of N experiments while in Section 2.1.3, we will be looking at it when the n experiments are chosen sequentially. Section 2.2 gives some extensions of those results to the cases where the parameter spaces of the k experiments are different and (or) no longer finite, and also to the case when the sample spaces are no longer finite.

2.1. A Characterization of the Class of Admissible Pairs in Finite Problems

2.1.1. For choosing one experiment

Consider the decision problem specified by a finite parameter space Θ which contains the true but unknown state of nature θ, a decision space D with generic element d, a nonnegative loss function $L(.,.)$ that satisfies the property that for any prior distribution λ on Θ, $\sum_{\theta} L(d,\theta)\lambda(\theta)$, as a function of d is uniquely minimized by a member of D, and a collection of k random variables $\{X_1,\ldots,X_k\}$ with finite sample spaces $\{\mathcal{X}_1,\ldots,\mathcal{X}_k\}$ and families $\{F_1,\ldots,F_k\}$ of possible probability functions where $F_i = \{f_\theta^i\colon \theta\in\Theta\}$ and F_i satisfies the assumption that for each $x_i \in \mathcal{X}_i$ there exists a $\theta \in \Theta$ such that $f_\theta^i(x_i) > 0$. Let δ_i $i = 1,\ldots,k$ denote a typical decision function (possibly randomized) from $\mathcal{X}_i$ to D with risk function $r_i(\delta_i;\theta)$. Let $\Gamma = \{\underset{\sim}{\gamma}\}$ be the class of all possible probability measures defined on $\{1,2,\ldots,k\}$ where $\underset{\sim}{\gamma} = (\gamma_1,\ldots,\gamma_k)$ and γ_i is the probability of observing X_i under $\underset{\sim}{\gamma}$. Let Γ^* be an arbitrary subset of Γ. (For example, suppose observing X_i costs c_i units then Γ^* could be $\{\underset{\sim}{\gamma}\colon \sum_{i=1}^{k} \gamma_i c_i \le c\}$ where c is some fixed cost such that $c_{i_1} < c_{i_2} < \ldots < c_{i_\ell} < c < c_{i_{\ell+1}} < \ldots < c_{i_k}$.) The problem is how to choose the pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ where $\underset{\sim}{\gamma} \in \Gamma^*$ and $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k)$.

Before stating the main results of this section, which characterizes the class of admissible pairs for this problem, we need the following notations and definitions:

For a pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$, the risk function is

$$r(\underset{\sim}{\gamma},\underset{\sim}{\delta};\theta) = \sum_{i=1}^{k} \gamma_i r_i(\delta_i;\theta)$$

and the Bayes risk against some prior distribution, say λ, is

$$R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda) = \sum_{i=1}^{k} \gamma_i R_i(\delta_i;\lambda)$$

where $R_i(\delta_i;\lambda)$ is the Bayes risk of δ_i against λ.

Let $\pi_i(x_i;\lambda) = \sum_{\theta} f_\theta^i(x_i)\lambda(\theta)$, $i = 1,\ldots,k$ be the marginal probability function of X_i under the prior λ. Two priors λ^i and λ^j $(i \neq j)$ are said to be orthogonal if $\Theta(\lambda^i) \cap \Theta(\lambda^j)$ is empty where $\Theta(\lambda^r) = \{\theta: \lambda^r(\theta) > 0\}$ $r = i,j$. For a set of priors $\lambda^1,\ldots,\lambda^m$, define the following sets associated with the i^{th} random variable:

$$\Lambda_i^1 = \{x_i: \pi_i(x_i;\lambda^1) > 0\}$$

and

$$\Lambda_i^r = \{x_i: x_i \not\in \bigcup_{j=1}^{r-1} \Lambda_i^j \quad \text{and} \quad \pi_i(x_i;\lambda^r) > 0\}$$

$$\text{for} \quad r = 2,\ldots,m\,.$$

Note that some of the Λ_i^r's might be empty and that the set associated with a particular prior depends on the other priors in the sequence and its place in the sequence.

Definition 2.1 (Meeden and Ghosh (1983)):

A decision rule δ_i, defined on X_i, is said to be stepwise Bayes against $\lambda^1,\ldots,\lambda^m$ if $\delta_i(x_i) = \delta_i^r(x_i)$ for all $x_i \in \Lambda_i^r$ for $r = 1,\ldots,m$ where δ_i^r is Bayes against λ^r.

From this definition, we notice that a stepwise Bayes rule is defined in terms of an ordered set of priors and a different ordering of those priors often results in a different stepwise Bayes rule. This definition also implies that a stepwise Bayes rule against $\lambda^1,\ldots,\lambda^m$ is necessarily a Bayes rule against λ^1, but it need not be Bayes against λ^r, $r = 2,\ldots,m$. Finally, we see that if $\lambda^1,\ldots,\lambda^m$ are such that $\bigcup_{r=1}^{m} \Theta(\lambda^r) = \Theta$ then the stepwise Bayes rule in this case is unique, however, the converse is not necessarily true.

Now, the following theorem provides a characterization of the class of admissible pairs for the problem.

Theorem 2.1:

Suppose that $\lambda^1,\ldots,\lambda^m$ is a set of mutually orthogonal prior distributions such that:

i) $\bigcup_{i=1}^{k} \Lambda_i^r$ is nonempty for $r = 1,2,\ldots,m$ (2.1)

ii) $$\bigcup_{r=1}^{m} \left(\bigcup_{i=1}^{k} \Lambda_i^r \right) = \bigcup_{i=1}^{k} X_i \qquad (2.2)$$

iii) δ_i is stepwise Bayes against $\lambda^1,\ldots,\lambda^m$ for the i^{th} problem $i = 1,\ldots,k$ (2.3)

Define the following sets:

$$\phi(\lambda^1) = \{\underset{\sim}{\gamma}: R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^1) = \inf_{\underset{\sim}{\gamma}\in\Gamma^*} R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^1)\}$$

$$\phi(\lambda^1,\lambda^2) = \{\underset{\sim}{\gamma}: R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^2) = \inf_{\underset{\sim}{\gamma}\in\phi(\lambda^1)} R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^2)\}$$

$$\vdots$$

$$\phi(\lambda^1,\ldots,\lambda^m) = \{\underset{\sim}{\gamma}: R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^m) = \inf_{\underset{\sim}{\gamma}\in\phi(\lambda^1,\ldots,\lambda^{m-1})} R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^m)\}$$

Then for any $\underset{\sim}{\gamma} \in \phi(\lambda^1,\ldots,\lambda^m)$, $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is admissible relative to Γ^* if and only if $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is admissible relative to $\phi(\lambda^1,\ldots,\lambda^m)$.

Before proving the theorem, we prove the following lemma.

<u>Lemma 2.1</u>:

Let $\lambda^1,\ldots,\lambda^m$ be a set of mutually orthogonal priors and let $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ be a pair such that δ_i is stepwise Bayes against $\lambda^1,\ldots,\lambda^m$ for the i^{th} problem $i = 1,\ldots,k$ and $\underset{\sim}{\gamma} \in \phi(\lambda^1,\ldots,\lambda^m)$. If $(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^o)$ is a pair which is at least as good as $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ then, $\underset{\sim}{\delta} = \underset{\sim}{\delta}^o$ and $\underset{\sim}{\gamma}^o \in \phi(\lambda^1,\ldots,\lambda^m)$ as well.

<u>Proof</u>:

Since $(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^o)$ is at least as good as $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$, then

$$r(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};\theta) \le r(\underset{\sim}{\gamma},\underset{\sim}{\delta};\theta) \qquad \forall\ \theta \in \Theta$$

and, hence,

$$R(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};\lambda^{r}) \le R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^{r}) \qquad \forall\ r = 1,\ldots,m\ . \tag{2.4}$$

On the other hand, for $m = 1$ we have

$$\begin{aligned} R(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};\lambda^{1}) &\ge R(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta};\lambda^{1}) \qquad (\because\ \underset{\sim}{\delta} \text{ is stepwise Bayes}) \\ &\ge R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^{1}) \qquad (\because\ \underset{\sim}{\gamma} \in \phi(\lambda^{1})) \end{aligned} \tag{2.5}$$

(2.4), (2.5) and the uniqueness assumption on the loss function imply that $\delta_i = \delta_i^o$ for $x_i \in \Lambda_i^1$ $i = 1,\ldots,k$ and $\underset{\sim}{\gamma}^o \in \phi(\lambda^1)$.

Assume for $m = n$ that $\delta_i = \delta_i^o$ for $x_i \in \bigcup_{r=1}^{n} \Lambda_i^r$ $i = 1,\ldots,k$ and $\underset{\sim}{\gamma}^o \in \phi(\lambda^1,\ldots,\lambda^n)$. Now, using this induction hypothesis we find that for $m = n+1$,

$$R(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};\lambda^{n+1}) \ge R(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta};\lambda^{n+1}) \ge R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^{n+1}) \tag{2.6}$$

(2.4), (2.6) and the uniqueness assumption on the loss function imply that $\delta_i = \delta_i^o$ on $\bigcup_{r=1}^{n+1} \Lambda_i^r$ $i = 1,\ldots,k$ and $\underset{\sim}{\gamma}^o \in \phi(\lambda^1,\ldots,\lambda^{n+1})$.

Note that the lemma is also true if $(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^o)$ dominates $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$.

Proof of Theorem 2.1:

Suppose $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is admissible relative to $\phi(\lambda^1,\ldots,\lambda^m)$. If $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is not admissible relative to Γ^*, then it is dominated by some other pair, say $(\underset{\sim}{\gamma}^{o}, \underset{\sim}{\delta}^{o})$, where $\underset{\sim}{\gamma}^{o} \in \Gamma^*$. By Lemma 2.1, $\underset{\sim}{\delta}^{o} = \underset{\sim}{\delta}$ and $\underset{\sim}{\gamma}^{o} \in \phi(\lambda^1,\ldots,\lambda^m)$, i.e., $(\underset{\sim}{\gamma}^{o}, \underset{\sim}{\delta})$ dominates $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ where both $\underset{\sim}{\gamma}$ and $\underset{\sim}{\gamma}^{o}$ belong to $\phi(\lambda^1,\ldots,\lambda^m)$ which is a contradiction.

Now, suppose $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ is admissible relative to Γ^*, but not admissible relative to $\phi(\lambda^1,\ldots,\lambda^m)$ then it is dominated by some other pair, say $(\underset{\sim}{\gamma}^*, \underset{\sim}{\delta}^*)$, with $\underset{\sim}{\gamma}^* \in \phi(\lambda^1,\ldots,\lambda^m)$ which is a contradiction and the proof is complete.

Remark 2.1:

It is obvious that if there exists r^* where $r^* = 1,\ldots,m$ such that $\phi(\lambda^1,\ldots,\lambda^{r^*})$ contains only one element, say $\underset{\sim}{\gamma}'$, then $(\underset{\sim}{\gamma}', \underset{\sim}{\delta})$ is admissible relative to Γ^*. On the other hand, if $\phi(\lambda^1,\ldots,\lambda^m)$ contains more than one element, then according to this theorem we have to show admissibility relative to $\phi(\lambda^1,\ldots,\lambda^m)$ in order to show admissibility relative to Γ^*. However, part (i) of the following Corollary shows that under an additional assumption, it is enough to have $\underset{\sim}{\gamma} \in \phi(\lambda^1,\ldots,\lambda^m)$ in order to have admissibility relative to Γ^*.

Corollary 2.1:

i) If $\phi(\lambda^1,\ldots,\lambda^m)$ contains more than one element and if for all $\underset{\sim}{\gamma} \neq \underset{\sim}{\gamma}' \in \phi(\lambda^1,\ldots,\lambda^m)$ (or $\in \Gamma^*$) neither $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ dominates $(\underset{\sim}{\gamma}', \underset{\sim}{\delta})$ nor vice versa, then for any $\underset{\sim}{\gamma} \in \phi(\lambda^1,\ldots,\lambda^m)$, $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ is admissible relative to Γ^*.

ii) If $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ is admissible relative to Γ^*, then there exists a set of mutually orthogonal priors such that (2.1), (2.2) and (2.3) are true and $\underset{\sim}{\gamma} \in \phi(\lambda^1,\ldots,\lambda^m)$.

Proof:

i) Suppose $(\underset{\sim}{\gamma}',\underset{\sim}{\delta}')$ where $\underset{\sim}{\gamma}' \in \Gamma^*$ dominates $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$, then by Lemma 2.1, $\underset{\sim}{\delta}' = \underset{\sim}{\delta}$ and $\underset{\sim}{\gamma}' \in \phi(\lambda^1,\ldots,\lambda^m)$, i.e., $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ dominates $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ where both $\underset{\sim}{\gamma}$ and $\underset{\sim}{\gamma}'$ belong to $\phi(\lambda^1,\ldots,\lambda^m)$ which is a contradiction.

ii) Since $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is admissible relative to Γ^*, then there exists a prior, say λ^1, against which $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is Bayes. Note that $R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^1) = \sum_{i=1}^{K} \gamma_i R_i(\delta_i;\lambda^1)$, i.e., δ_i restricted to Λ_i^1 is Bayes for $i = 1,\ldots,k$. Now, it must be the case that $\underset{\sim}{\gamma} \in \phi(\lambda^1)$ otherwise $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is not Bayes against λ^1. Let $\phi^*(\lambda^1) = \{(\underset{\sim}{\gamma},\underset{\sim}{\delta}^o)\colon \delta_i^o = \delta_i$ on Λ_i^1 for $i = 1,\ldots,k\}$, i.e., $\phi^*(\lambda^1)$ is the class of all pairs which are Bayes against λ^1. Note that $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ belongs to $\phi^*(\lambda^1)$. If $\phi^*(\lambda^1)$ consists of only one pair, namely $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$, then the proof is complete. So suppose this is not the case and consider the restricted problem with $\theta \notin \Theta(\lambda^1)$ and the risk set of $\phi^*(\lambda^1)$. For this problem $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is admissible with respect to $\phi^*(\lambda^1)$, i.e., there does not exist any member in $\phi^*(\lambda^1)$ that dominates $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ otherwise $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is inadmissible for the original problem as well. Hence, there exists a prior, say λ^2, against which $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is Bayes. As before, δ_i restricted to $\Lambda_i^1 \cup \Lambda_i^2$ is Bayes for $i = 1,\ldots,K$ and $\underset{\sim}{\gamma} \in \phi(\lambda^1,\lambda^2)$. Let $\phi^*(\lambda^1,\lambda^2) = \{(\underset{\sim}{\gamma},\underset{\sim}{\delta}^o)\colon \delta_i^o = \delta_i$ on $\Lambda_i^1 \cup \Lambda_i^2$ for $i = 1,\ldots,k\}$, i.e., $\phi^*(\lambda^1,\lambda^2)$ is the class of all pairs which are

Bayes within $\phi^*(\lambda^1)$ against λ^2. Continue this way until we get a set of mutually orthogonal priors, say $\lambda^1,\ldots,\lambda^m$, such that $\phi^*(\lambda^1,\ldots,\lambda^m)$ consists of only one element, namely $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$. Remove from this set of priors all priors λ^r under which $\bigcup_{i=1}^{k} \Lambda_i^r$ is empty. Note that m is finite since Θ is finite and the proof is complete.

Remark 2.2:

If the set of priors is such that $\Theta = \bigcup_{r=1}^{m} \Theta(\lambda^r)$, then it is easy to see that part (i) of Corollary 2.1 is true without the assumption that "for all $\underset{\sim}{\gamma} \neq \underset{\sim}{\gamma}' \in \phi(\lambda^1,\ldots,\lambda^m)$ (or Γ^*) neither $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ dominates $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ nor vice versa." This alternative assumption will be used later in the applications.

Note that without any of the above two assumptions, the class $\phi(\lambda^1,\ldots,\lambda^m)$ might contain inadmissible solutions. To see that we consider the following example:

Let E_1 and E_2 be two experiments with corresponding random variables X_1 and X_2 and

$$r(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta};\theta) = r(\underset{\sim}{\gamma}',\delta;\theta) \qquad \forall\ \theta \neq \theta_o$$

and

$$r(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta};\theta_o) < r(\underset{\sim}{\gamma}',\delta;\theta_o)$$

where $\underset{\sim}{\gamma}^o = (1 \quad 0)$, $\underset{\sim}{\gamma}' = (\gamma_1' \quad \gamma_2')$ with $0 \leq \gamma_1' < 1$ and $\underset{\sim}{\delta}$ is unique stepwise Bayes against some sequence of priors. Suppose that observing E_i costs c_i units $i = 1,2$ where $c_2 < c < c_1$ and we want to choose between these two experiments such that the expected cost is not more than c. In this case, $\Gamma^* = \{\underset{\sim}{\gamma}: \sum_{i=1}^{2} \gamma_i c_i \leq c\}$. Now, assume that none of the priors puts positive mass on θ_o. In this case, $R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda^r)$ is the same for any $\underset{\sim}{\gamma} \in \Gamma^*$ and for all $r = 1,\ldots,m$, i.e., $\phi(\lambda^1,\ldots,\lambda^m) = \Gamma^*$. Hence, $\underset{\sim}{\gamma}' = (0 \quad 1) \in \phi(\lambda^1,\ldots,\lambda^m)$ although $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ is inadmissible. In this case, it is easy to see that the admissible pair is $(\underset{\sim}{\gamma}^*,\underset{\sim}{\delta})$ where $\underset{\sim}{\gamma}^*$ is such that $\sum_{i=1}^{2} \gamma_i^* c_i = c$.

On the other hand, with either of the two assumptions mentioned in Remark 2.2, if the problem has no admissible solutions then $\phi(\lambda^1,\ldots,\lambda^m)$ will be empty. But without any of these assumptions there could be some cases where the problem has no admissible solutions and $\phi(\lambda^1,\ldots,\lambda^m)$ is nonempty. This can be shown using the example given in the previous paragraph but with Γ^* as $\Gamma^* = \{\underset{\sim}{\gamma}: \sum_{i=1}^{2} \gamma_i c_i < c\}$. In this case, there are no admissible solutions. However, $\phi(\lambda^1,\ldots,\lambda^m) = \Gamma^*$.

Remark 2.3:

In Meeden and Ghosh (1983), it was assumed that "for all $i \neq i' = 1,\ldots,k$ neither $r_i(\delta_i;\theta)$ dominates $r_{i'}(\delta_{i'};\theta)$ nor vice versa." For $k = 2$, it is easy to see that this assumption implies the assumption that "for all $\underset{\sim}{\gamma} \neq \underset{\sim}{\gamma}' \in \phi(\lambda^1,\ldots,\lambda^m)$ (or $\in \Gamma^*$) neither $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ dominates $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ nor vice versa," (i.e., the assumption given in part (i) of Corollary 2.1) and, hence, it is enough to use the first assumption instead of the later one. However, the following example shows that this is not the case for $k > 2$:

Let X_i, $i = 1,2,3$ be three random variables with $\mathcal{X}_i = \{0,1\}$ and $\Theta = \{0, 1/3, 2/3, 1\}$. Let $f_0^i(0) = f_1^i(1) = 1$ for $i = 1,2,3$, $f_{1/3}^1(0) = f_{1/3}^2(1) = 1/4$, $f_{2/3}^1(0) = f_{2/3}^2(1) = f_{1/3}^3(0) = 1/3$ and $f_{2/3}^3(0) = 1/2$. Let λ be the prior that assigns its mass to 0 and 1. It is easy to see, using squared error loss, that the unique Bayes estimator δ_i for X_i is $\delta_i(0) = 0 = 1-\delta_i(1)$ for $i = 1,2,3$ and the risk vectors are $(0, 13/36, 8/36, 0)$, $(0, 7/36, 12/36, 0)$ and $(0, 12/36, 10/36, 0)$. It is obvious that for all $i \neq i' = 1,2,3$ neither $r_i(\delta_i;\theta)$ dominates $r_{i'}(\delta_{i'};\theta)$ nor vice versa. Now, consider the pairs $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ and $(\underset{\sim}{\gamma}'',\underset{\sim}{\delta})$ where $\underset{\sim}{\gamma}' = (1/13 \quad 10/13 \quad 2/13)$ and $\underset{\sim}{\gamma}'' = (0 \quad 9/13 \quad 4/13)$. It is easy to show that $r(\underset{\sim}{\gamma}',\underset{\sim}{\delta};\theta) = r(\underset{\sim}{\gamma}'',\underset{\sim}{\delta};\theta) = 0$ for $\theta = 0,1$ and $r(\underset{\sim}{\gamma}',\underset{\sim}{\delta};2/3) = r(\underset{\sim}{\gamma}'',\underset{\sim}{\delta};2/3) = 148$. But $r(\underset{\sim}{\gamma}',\underset{\sim}{\delta};1/3) = 107 < 111 = r(\underset{\sim}{\gamma}'',\underset{\sim}{\delta};1/3)$, i.e., $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ dominates $(\underset{\sim}{\gamma}'',\underset{\sim}{\delta})$. From this, we see that the assumption given by Meeden and Ghosh (1983) is satisfied while the assumption given in part (i) of Corollary 2.1 is not. Now, it is clear that $R(\underset{\sim}{\gamma},\underset{\sim}{\delta};\lambda) = 0$ for any $\underset{\sim}{\gamma}$ and, hence, $\phi(\lambda) = \Gamma$. In particular, $\gamma'' \in \phi(\lambda)$ although $(\underset{\sim}{\gamma}'',\underset{\sim}{\delta})$ is inadmissible. This shows that the assumption that "for all $i \neq i' = 1,2,3$ neither $r_i(\delta_i;\theta)$ dominates $r_{i'}(\delta_{i'};\theta)$ nor vice versa" is not enough to ensure that for every member in $\phi(\lambda)$ the corresponding pair is admissible.

<u>Remark 2.4</u>:

We have assumed that the loss function is such that the prior risk is uniquely minimized by a member of D. A case in which this assumption is satisfied is when the loss function is strictly convex (DeGroot and Rao (1963)). An example of such loss function is the squared error loss function.

2.1.2. For choosing nonsequentially more than one experiment

Through the previous section we have been looking at the problem of characterizing the admissible pairs in the case when the statistician is interested in making his decision by choosing one experiment out of k experiments. However, there might be some cases where the statistician may want to make his decision by choosing nonsequentially a subset of the k experiments rather than a single one. For this reason, we devote this section to show how the admissible pairs in this problem can be characterized using the results of the former section.

Let Θ, D and $L(.,.)$ be as specified in the previous section. Let $X^* = \{X_1^*,\ldots,X_N^*\}$ be a collection of N independent random variables with finite sample spaces $\{\mathcal{X}_1^*,\ldots,\mathcal{X}_N^*\}$ and corresponding families of possible probability functions $\{H_1,\ldots,H_N\}$ where $H_i = \{h_\theta^i: \theta \in \Theta\}$ $i = 1,\ldots,N$. Assume that H_i satisfies the property that for each $x_i^* \in \mathcal{X}_i^*$ there exists a $\theta \in \Theta$ such that $h_\theta^i(x_i^*) > 0$. Now, the interest is to make some decision about θ after choosing nonsequentially a subset of size n from X^*.

This problem can be reformulated so that it becomes of the form of the decision problem given in the previous section. For instance, since the interest is to choose a subset of size n from X^* then there are $\binom{N}{n}$ such subsets. Set $k = \binom{N}{n}$ and let $X = \{\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_k\}$, i.e., X is the set of k random vectors of length n generated from X^*. If a member of X, say $\underset{\sim}{X}_j$, comes from choosing $X^*_{i_1}, X^*_{i_2}, \ldots, X^*_{i_n}$ where without loss of generality we assume $i_1 < i_2 < \ldots < i_n$, then the sample space of $\underset{\sim}{X}_j$, say $\mathcal{X}_j$, is $\mathcal{X}_j = \{\underset{\sim}{x}_j = (x_{j_1}, x_{j_2}, \ldots, x_{j_n}\}$ where $x_{j_\ell} \in \mathcal{X}^*_{i_\ell}$: for each $\underset{\sim}{x}_j$ $\exists$ a $\theta \in \Theta$ such that $f_\theta(\underset{\sim}{x}_j) = \prod_{\ell=1}^{n} h_\theta^{j_\ell}(x^*_{j_\ell}) > 0\}$.

Letting δ_j $j = 1,\ldots,k$ denote a typical decision function (possibly randomized) from $\mathcal{X}_j$ to D with risk function $r_j(\delta_j;\theta)$ and $\Gamma^* = \{\underset{\sim}{\gamma}\}$ be a class of probability measures defined on $\{1,\ldots,k\}$ we see that the problem of choosing an admissible pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ relative to Γ^* in this case is of the type considered in the previous section and, hence, all the results given in that section are applicable here.

If we call $\underset{\sim}{\gamma}$ a design and we take $\Gamma^* = \Gamma$ then this version gives admissibility results relative to the class of designs of fixed sample size n.

Remark 2.5:

It is easy now to see that the above ideas can be extended to the more general case when the statistician can make his decision by choosing any subset of any size from X^*. In this case, X^* generates the set of subsets $X = \{\{X_1\}, \{X_2\}, \ldots, \{X_N\}\}$ where $\{X_i\}$ is the collection of subsets of size i from X^*. Note X has $\sum_{\ell=1}^{N} \binom{N}{\ell}$

subsets. Set $k = \sum_{\ell=1}^{N} \binom{N}{\ell}$ and let δ_j $j = 1,\ldots,k$ be a decision function (possibly randomized) from X_j, the sample space of the j^{th} subset, to D. Letting $\Gamma^* = \{\underset{\sim}{\gamma}\}$ be a class of discrete probability measures defined on $\{1,\ldots,k\}$ we see that the admissible pairs $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ relative to Γ^* can be characterized using the procedure given in Section 2.1.1.

As above, if we call $\underset{\sim}{\gamma}$ a design and we take $\Gamma^* = \{\underset{\sim}{\gamma}: \sum_{j=1}^{k} \gamma_j n_j = n\}$ where n_j is the size of the j^{th} subset, then any admissibility results obtained this way will be relative to the class of designs of expected sample size n.

Note that both the classes of designs of fixed sample size n and of expected sample size n are common classes in finite population sampling where samplers are usually interested in proving admissibility relative to either one of them. In Chapter 3, we will be using the ideas of this section to give some admissibility results relative to those classes of designs.

2.1.3. For choosing sequentially more than one experiment

In this section, we consider the problem of characterizing the admissible pairs when choosing, contrary to the previous section, sequentially a subset of experiments.

Consider a decision problem with Θ, D and $L(\cdot,\cdot)$ as specified in Section 2.1.1. Let $X^* = \{X_1^*,\ldots,X_N^*\}$ be a collection of N independent random variables (that correspond to N experiments) with finite sample spaces $\{\mathcal{X}_1^*,\ldots,\mathcal{X}_N^*\}$ and corresponding families of possible probability functions $\{H_1,\ldots,H_N\}$ where $H_i = \{h^i_\theta\colon \theta \in \Theta\}$ $i = 1,\ldots,N$. Assume that H_i satisfies the property that for each $x^*_i \in X^*_i$ there exists a $\theta \in \Theta$ such that $h^i_\theta(x^*_i) > 0$.

The interest is to estimate some parametric function, say $\tau(\theta)$, based on choosing sequentially a subset of size $n (n < N)$ from X^*.

This problem can be viewed as follows: Let $\mathcal{Q}$ denote the class of all probability measures, $q(x^*)$, that chooses sequentially a subset of size n from X^*. Since each X^*_i is finite, $\mathcal{Q}$ contains only a finite number of elements, say k, i.e.

$$\mathcal{Q} = \{q_j(x^*)\colon j = 1,2,\ldots,k\}.$$

A typical element in $\mathcal{Q}$ can be written as

$$q_j(x^*) = p_j(i_1)p_j(i_2|x^*_{i_1})p_j(i_3|x^*_{i_1},x^*_{i_2}) \ldots p_j(i_n|x^*_{i_1},\ldots,x^*_{i_{n-1}})$$

where $p_j(i_1)$ is the probability, under $q_j(\cdot)$, of choosing a member of X^* to be observed first and $p_j(i_2|x^*_{i_1})$ is the probability, under $q_j(\cdot)$, of choosing a member from X^* to be observed second given that $X^*_{i_1} = x^*_{i_1}$ and so on.

Let $\mathcal{X}$ and $\mathcal{X}_j$ be defined as follows:

$$\mathcal{X} = \{\underset{\sim}{x} = (x^*_{i_1},\ldots,x^*_{i_n})\colon \text{ For each } \underset{\sim}{x} \ \exists \text{ a } \theta \in \Theta \ni h_\theta(\underset{\sim}{x}) = \prod_{j=1}^{n} h_\theta^{i_j}(x^*_{i_j}) > 0\}$$

$$\mathcal{X}_j = \{\underset{\sim}{x} \in \mathcal{X}\colon \ q_j(\underset{\sim}{x}) > 0\} \quad j = 1,2,\ldots,k,$$

i.e., X is the set of all possible vectors of length n chosen from X^* such that each $\underset{\sim}{x} \in X$ can be observed under at least one $\theta \in \Theta$. While X_j is the subset of X that receives positive mass under $q_j(\cdot)$. Now, taking X_j to be the sample space of a random vector, say $\underset{\sim}{X}_j$, we see that the probability function, say $f_\theta^j(\cdot)$, is

$$f_\theta^j(\underset{\sim}{x}) = h_\theta(\underset{\sim}{x}) q_j(\underset{\sim}{x}) \quad \text{for} \quad \underset{\sim}{x} \in X_j.$$

From now on we will call $q_j(\cdot)$ a sequential design.

It is clear now that each sequential design $q_j(\cdot)$ defines a finite subset X_j of X. Those finite subsets $X_1,\ldots,X_k$ can be viewed as sample spaces of random vectors, say $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_k$, with families of possible probability functions $\{F_1,\ldots,F_k\}$ where $F_j = \{f_\theta^j\colon \theta \in \Theta\}$ $j = 1,\ldots,k$ satisfies the assumption that for each $\underset{\sim}{x} \in X_j$ there exists a $\theta \in \Theta$ such that $f_\theta^j(\underset{\sim}{x}) > 0$. Now, letting δ_j denote a typical decision function (possibly randomized) from X_j to D with risk function $r_j(\delta_j;\theta)$, we see that the problem of choosing a sequential design along with a decision rule is of the form of the problem of choosing between experiments given in Meeden and Ghosh (1983) (and also of the type of the problem considered in Section 2.1.1 with $\Gamma^* = \Gamma$). For the purpose of completeness, and utilizing the theorem given in Meeden and Ghosh (1983), we give the following theorem which provides a characterization of the admissible pairs for this problem:

Theorem 2.2:

a) Let $\lambda^1,\ldots,\lambda^m$ be a set of mutually orthogonal prior distributions such that

$$\text{i)} \quad \bigcup_{j=1}^{k} \Lambda_j^r \text{ is nonempty for } r = 1,2,\ldots,m \tag{2.7}$$

$$\text{ii)} \quad \bigcup_{r=1}^{m} \left(\bigcup_{j=1}^{k} \Lambda_j^r \right) = \bigcup_{j=1}^{k} X_j \tag{2.8}$$

iii) δ_j is stepwise Bayes against $\lambda^1,\ldots,\lambda^m$ for the $\underset{\sim}{X}_j$ problem $j = 1,\ldots,k$ (2.9)

Define the following sets:

$$\varphi(\lambda^1) = \{j\colon R_j(\delta_j;\lambda^1) = \inf_{1\le j\le k} R_j(\delta_j;\lambda^1)\}$$

$$\varphi(\lambda^1,\lambda^2) = \{j\colon R_j(\delta_j;\lambda^2) = \inf_{j\in\varphi(\lambda^1)} R_j(\delta_j;\lambda^2)\}$$

$$\vdots$$

$$\varphi(\lambda^1,\ldots,\lambda^m) = \{j\colon R_j(\delta_j;\lambda^m) = \inf_{j\in\varphi(\lambda^1,\ldots,\lambda^{m-1})} R_j(\delta_j;\lambda^m)\}.$$

Then,

a.1) If there exists r^* such that $\varphi(\lambda^1,\ldots,\lambda^{r^*})$ consists of only one element, say j^*, then (q_{j^*}, δ_{j^*}) is admissible relative to Q.

a.2) If $\varphi(\lambda^1,\ldots,\lambda^m)$ contains more than one element and if $\bigcup_{r=1}^{m} \Theta(\lambda^r) = \Theta$ then for any $j \in \varphi(\lambda^1,\ldots,\lambda^m)$, (q_j, δ_j) is admissible relative to Q. (Also, any random choice between those pairs will result in an admissible pair.)

b) If $(q_{j'}, \delta_{j'})$ is admissible relative to Q then there exists a set of mutually orthogonal priors, say $\lambda^1,\ldots,\lambda^m$, such that (2.7), (2.8) and (2.9) are true and $j' \in \varphi(\lambda^1,\ldots,\lambda^m)$.

Remark 2.6:

Note that part a.2) is true if we replace the assumption "$\bigcup_{r=1}^{m} \Theta(\lambda^r) = \Theta$" by the assumption that "for any two probability distributions $\underset{\sim}{\gamma}$ and $\underset{\sim}{\gamma}'$ defined on $\varphi(\lambda^1,\ldots,\lambda^m)$, neither $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ dominates $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ nor vice versa where $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_k)$".

The term "pair" used here to denote (q_j, δ_j) say, doesn't have the same meaning as that used in previous sections. For instance, the notation (q_j, δ_j) is used to indicate experiment $\underset{\sim}{X}_j$ that is defined by the sequential design q_j and δ_j is the decision rule to be used in connection with $\underset{\sim}{X}_j$. For such a pair, the risk function is

$$r(q_j,\delta_j;\theta) = r(\delta_j;\theta)$$

and the Bayes risk against some prior, say λ, is

$$R(q_j,\delta_j;\lambda) = R_j(\delta_j;\lambda).$$

From the way the sets $\varphi(\lambda^1)$, $\varphi(\lambda^1,\lambda^2)$, $\ldots$, $\varphi(\lambda^1,\ldots,\lambda^m)$ are constructed, we see that for any $j^* \in \varphi(\lambda^1,\ldots,\lambda^m)$, the corresponding q_{j^*} is just the sequential design that gives, at each stage, the minimum Bayes risk among all other sequential designs. This design,

q_{j*}, is in fact the sequential design that is determined by backward induction. Utilizing one prior distribution, a description of backward induction in sequential problems that deals with stopping rules (i.e., with sequential problems where the number of observations is not fixed in advance) is given in DeGroot (1970). A slight modification to this description yields a backward induction technique that fits the frame of the sequential problem considered in this section, namely, the sequential problem where the number of observations is fixed in advance. This modification, also, may utilize a set of mutually orthogonal prior distributions.

2.2. Extension of the Results of Section 2.1

2.2.1. To the case of different parameter spaces

Throughout Section 2.1.1, we have assumed that for any X_i $i = 1,\dots,k$ the parameter space is Θ. Now, suppose this is not the case, i.e., suppose that X_i has a parameter space Θ_i then by taking the union of those Θ_i's that are different to be a common parameter space, say Θ, for the X_i's $i = 1,\dots,k$ and defining for X_i a new sample space $\mathcal{X}_i' = \mathcal{X}_i \cup \{a_i\}$ along with a new probability function $\mathfrak{f}_\theta^i(x_i')$ where

$$
\begin{aligned}
\mathfrak{f}_\theta^i(x_i') = f_\theta^i(x_i) \quad & \text{for} \quad \theta \in \Theta_i \ \text{ and } \ x_i' \neq a_i \\
0 \quad & \text{for} \quad \theta \in \Theta_i \ \text{ and } \ x_i' = a_i \\
0 \quad & \text{for} \quad \theta \notin \Theta_i \ \text{ and } \ x_i' \neq a_i \\
1 \quad & \text{for} \quad \theta \notin \Theta_i \ \text{ and } \ x_i' = a_i \,.
\end{aligned}
$$

We see that all the assumptions required in Section 2.1.1 are satisfied in this case and, hence, all the results given in that section remain true.

This argument is also applicable to Sections 2.1.2 and 2.1.3. However, it might not be useful in the cases when the Θ_i's are much different from each other.

2.2.2. To the case of nonfinite problems

In Section 2.1, we have assumed that the sample spaces, X_i's $i = 1,\ldots,k$, are finite. However, the results of that section remain true in the case when the sample spaces, X_i's, are countable provided that each decision rule δ has a finite risk.

Also, in Section 2.1, the parameter space is taken to be a finite set. However, as we see from the proofs, Theorem 2.1, and part (i) of Corollary 2.1 are independent of that assumption. That is, if the parameter space is not finite and it is easy to define a full sequence of orthogonal priors on that parameter space then both the theorem and that part of the corollary are applicable. However, in some cases as we will see in the sampling problem, it is not easy to define a sequence of priors on a parameter space that is not finite. For such cases, the following notion of "finite admissibility" introduced by Meeden and Ghosh (1982) can be used to prove admissibility:

Definition 2.2:

An estimator δ (or a pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$) is said to be finitely admissible (or finitely admissible relative to Γ^*) if given any parameter point $\theta_o \in \Theta$, there exists a finite subset Θ_o of Θ containing θ_o such that if θ is assumed to belong to Θ_o then δ (or $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$) is admissible (admissible relative to Γ^*).

As shown by Meeden and Ghosh (1982), every finitely admissible estimator (or pair) is admissible.

According to this notion of finite admissibility, in order to prove admissibility when the parameter space is no longer finite we need to prove admissibility on every finite subset of the parameter space. This might seem hard to do, however, as we will see in the next chapter this is very easy when choosing properly a finite subset of the parameter space that is rich enough so that when admissibility is proved on it, it insures admissibility on every finite subset of the parameter space. In fact, as we will see in the next chapter, Meeden and Ghosh have introduced such a subset to prove admissibility in finite population sampling.

3. APPLICATIONS

3.1. In Finite Population Sampling

3.1.1. Some preliminaries

Consider a finite population with units labeled $1, 2, \ldots, N$. Let y_i be the value of a single characteristic attached to the i^{th} unit. The vector $y = (y_1, \ldots, y_N)$ is the unknown state of nature and is assumed to belong to $\Theta = \mathbb{R}^N$, the N dimensional Euclidean space. A subset $s = \{i_1, \ldots, i_{n(s)}\}$ of $\{1, \ldots, N\}$ is called a sample of size $n(s)$. A discrete probability measure, p, defined on the set S of all possible samples from this population is called a design. Suppose that for estimating some real valued function, say $\tau(y)$, with squared error loss, one uses an estimator, say $e(s,y)$ ($e(s,y)$, depends on y only through $y(s) = (y_{i_1}, \ldots, y_{i_{n(s)}})$) along with a design p then $(p, e(s,y))$ is a typical decision strategy with risk function

$$
\begin{aligned}
r(p,e;y) &= \sum_{s \in S} [e(s,y) - \tau(y)]^2 \, p(s) \\
&= \sum_{s \in S} r_s(e(s,y);y) \, p(s) \,. \qquad (3.1)
\end{aligned}
$$

Now, we restate the definitions of admissibility in the frame work of finite population sampling.

<u>Definition 3.1</u>:

An estimator e is said to be admissible when using a design

p if there does not exist any other estimator e' with $r(p,e';y) \leq r(p,e;y)$, for all $y \in \Theta$ with strict inequality for some $y \in \Theta$.

Definition 3.2:

A pair (p,e) is said to be uniformly admissible relative to some class of designs, say $\mathcal{P}$, if $p \in \mathcal{P}$ and there does not exist any other pair (p',e'), with $p' \in \mathcal{P}$ such that $r(p',e';y) \leq r(p,e;y)$ for all $y \in \Theta$ with strict inequality for some $y \in \Theta$.

By considering the set of all possible samples from a given population to be the set of experiments available to a statistician, we see that the problem of choosing a uniformly admissible pair (p,e) relative to some class of designs is of the type considered in Chapter 2. Since the parameter space here is not finite, then the extension given in Section 2.2.2, which is based on the notion of finite admissibility, will be used. As we mentioned in Section 2.2.2, proving finite (uniform) admissibility will be easy if we choose properly a finite subset of the parameter space. In fact, Meeden and Ghosh (1983) introduced such a subset as follows: For any point $y^o \in \Theta$ containing distinct values $\alpha_1,\ldots,\alpha_r$ they took the finite subset to be $\bar{\Theta}(\alpha_1,\ldots,\alpha_r) = \{y\colon y_i = \alpha_j$ for some $j = 1,\ldots,r;$ for all $i = 1,\ldots,N\}$. It is obvious that $\bar{\Theta}(\alpha_1,\ldots,\alpha_r)$ is a finite subset of Θ and it contains y^o. Moreover, it is clear from the way $\bar{\Theta}(\alpha_1,\ldots,\alpha_r)$ is chosen that if an estimator (pair) is shown to be admissible (uniformly admissible) when considering $\bar{\Theta}(\alpha_1,\ldots,\alpha_r)$ then this result is also true when considering any other finite subset of Θ which implies that the estimator (pair) is finitely (uniformly) admissible. Using these kind of subsets along with the results given in Chapter 2, we will be giving some finite (uniform) admissibility results in finite population sampling. In particular, for estimating the population total, the finite uniform admissibility (and, hence, uniform admissibility) relative to the class of designs of expected sample size less than or equal to n of some different strategies is demonstrated in Section 3.1.2. While in Section 3.1.3, a finitely admissible (and, hence, an admissible) estimator of the population counterpart of a U-statistic is constructed following the line of argument given in Ghosh and Meeden (1982). In fact, this estimator turns out to be a multiple of the U-statistic.

3.1.2. Uniform admissibility when estimating the population total

For estimating the population total, $\sum_{i=1}^{N} y_i$, Basu (1971) has proposed the following estimator:

$$e_1(s,y) = \sum_{i\in s} y_i + (1/n(s))[\sum_{i\in s} (y_i/m_i)][\sum_{i\notin s} m_i] \tag{3.2}$$

where $m = (m_1,\ldots,m_N)$ is a vector of positive prior guesses for the vector $y = (y_1,\ldots,y_N)$. His motivation for this estimator is as follows: Suppose that before observing the sample, the statistician is willing to make a prior guess m_i for the value y_i $i = 1,\ldots,N$. After the sample s is observed the ratios y_i/m_i's, $i\in s$ become known. If these ratios are approximately equal, then the unobserved ratios will probably take on similar values as well. This suggests that given the sample one could assume that any unobserved ratio takes on the value of an observed ratio with probability $1/n(s)$. Therefore, given the sample, the expected value of any y_{i*}, $i^* \notin s$ is $(1/n(s))[m_{i*} \sum_{i\in s} (y_i/m_i)]$ and, hence, we have the estimator defined in (3.2).

Meeden and Ghosh (1983) have shown that $e_1(s,y)$ is admissible under any design. Now, the following theorem identifies some designs so that when used with $e_1(s,y)$ then the pair is uniformly admissible relative to the class of designs of expected sample size less than or equal to n. Before stating the theorem we need to introduce the following notations:

$\mathcal{P}_1 = \{p: \sum_{s\in S} n(s)p(s) \leq n\}$ i.e. $\mathcal{P}_1$ is the class of designs of expected sample size less than or equal to n $(n \leq N)$.

$S^* = \{s: s\in S_n$ and $\sum_{i\in s} m_i = \max_{s'\in S_n} \sum_{i\in s'} m_i\}$ where S_n is the set of samples of size n.

$\mathcal{P}_1^* = \{p: p\in \mathcal{P}_1$ and $p(s) = 0$ if $s\notin S^*\}$.

Theorem 3.1:

For estimating the population total with squared error loss, the pair $(p,e_1(s,y))$ where $p \in \mathcal{P}_1^*$ is uniformly admissible relative to $\mathcal{P}_1$.

Proof:

For any point $y^o \in \Theta$ containing distinct values $\alpha_1,\ldots,\alpha_r$ $(r \leq N)$

of the ratios y_i/m_i's let

$$\bar{\Theta}_m(\alpha_1,\ldots,\alpha_r) = \{y: y_i/m_i = \alpha_j \text{ for some } j = 1,\ldots,r; \text{ for all } i = 1,\ldots,N\}$$

and

$$\bar{\bar{\Theta}}_m(\alpha_1,\ldots,\alpha_r) = \{y: y_i/m_i = \alpha_j \text{ for some } j = 1,\ldots,r; \text{ for all } i = 1,\ldots,N \text{ and each } \alpha_j \text{ appears at least once for } j = 1,\ldots,r\}.$$

Now, define the set of priors $\lambda^1,\ldots,\lambda^r$ on $\bar{\Theta}_m(\alpha_1,\ldots,\alpha_r)$ in the following manner: λ^1 puts mass $\frac{1}{r}$ on each point in the set $\bigcup_{j=1}^{r} \bar{\bar{\Theta}}_m(\alpha_j)$ and λ^ℓ, $\ell = 2,\ldots,r$ is defined on the set $\bigcup_{j_1 < j_2 < \ldots < j_\ell} \bar{\bar{\Theta}}_m(\alpha_{j_1}, \alpha_{j_2}, \ldots, \alpha_{j_\ell})$ as follows:

$$\lambda^\ell(y) = \int_0^1 \cdots \int_0^1 \prod_{k=1}^{\ell} \nu_k^{w_y(j_k)-1} \prod_{k=1}^{\ell} d\nu_k$$

where $w_y(j_k)$ is the number of (y_i/m_i)'s in y which are equal to α_{j_k}. Note that $w_y(j_k) \geq 1$ for any j_k and $\sum_{k=1}^{r} w_y(j_k) = N$.

Meeden and Ghosh (1983) have shown that for any sample s, $e_1(s,y)$ is unique stepwise Bayes against this set of priors when the parameter space is $\bar{\Theta}_m(\alpha_1,\ldots,\alpha_r)$. Therefore, according to remark 2.2, in order to prove the above theorem it suffices to show that the class of designs $\phi(\lambda^1,\ldots,\lambda^r) = P_1^*$.

Letting $z_i = y_i/m_i$ $i = 1,\ldots,N$, the risk of a pair (p,e_1) where $p \in P_1$ is

$$r(p,e_1;y) = \sum_{s\in S} p(s)[n^{-1}(s)(\sum_{i\in s} z_i)(\sum_{i\notin s} m_i) - (\sum_{i\notin s} z_i m_i)]^2$$

$$= \sum_{s\in S} p(s)[\sum_{i=1}^{N} a_{i,s} z_i]^2$$

where $a_{i,s} = n^{-1}(s)(\sum_{i\notin s} m_i)$ for $i \in s$ and $a_{i,s} = -m_i$ for $i \notin s$. It is obvious that the Bayes risk of (p,e_1) under λ^1 is zero. Now, under any λ^ℓ, $\ell = 2,\ldots,r$ the ratios y_i/m_i's are finitely exchangeable. Therefore, the Bayes risk of (p,e_1) under λ^ℓ is

$$R(p,e_1;\lambda^\ell) = \sum_{s\in S} p(s)[(\sum_{i=1}^{N} a_{i,s}^2)\ E(z_1^2) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i\neq j}}^{N} a_{i,s} a_{j,s} E(z_1 z_2)]$$

where the expectations are taken with respect to the marginal priors under λ^ℓ. Since $\sum_{i=1}^{N} a_{i,s} = 0$ then,

$$R(p,e_1;\lambda^\ell) = \sum_{s\in S} p(s)[\sum_{i=1}^{N} a_{i,s}^2 \{E(z_1^2) - E(z_1 z_2)\}].$$

By the Schwarz inequality, we have $E(z_1^2) - E(z_1 z_2) \geq 0$. Hence,

$$\inf_{p\in P_1} R(p,e_1;\lambda^\ell) = [E(z_1^2) - E(z_1 z_2)] \inf_{p\in P_1} \sum_{s\in S} p(s)(\sum_{i=1}^{N} a_{i,s}^2)$$

$$= [E(z_1^2) - E(z_1 z_2)] \inf_{p\in P_1} \sum_{s\in S} p(s)[n^{-1}(s)(\sum_{i\notin s} m_i)^2 + \sum_{i\notin s} m_i^2] \quad (3.3)$$

By ordering the m_i's such that $m_{(1)} \geq m_{(2)} \geq \cdots \geq m_{(N)}$ and letting $\bar{S}_{n(s)} = \{s\colon s \in S_{n(s)}$ and $\sum_{i\in s} m_i = \sum_{j=1}^{n(s)} m_{(j)}\}$ where $S_{n(s)}$ is the set of samples of size $n(s)$, $\bar{S} = \bigcup_{n(s)=1}^{N} \bar{S}_{n(s)}$ and $\bar{P} = \{p\colon p \in P_1$ and $p(s) = 0$ if $s \notin \bar{S}\}$ we note that the set of designs that gives the infimum of (3.3) is a subset of $\bar{P}$. Therefore,

$$\inf_{p\in P_1} R(p,e_1;\lambda^{\ell}) = [E(z_1^2)-E(z_1z_2)]\inf_{p\in\bar{P}} \sum_{s\in\bar{S}} p(s)[n^{-1}(s)(\sum_{i\notin s} m_i)^2+\sum_{i\notin s} m_i^2]$$

$$= [E(z_1^2)-E(z_1z_2)] \inf_{p\in\bar{P}} \sum_{i=1}^{N} p(i)\Psi(i)$$

where $p(i)$ is the probability, under the design p, of selecting the sample of size i that has the largest m_i's and $\Psi(i)$ is

$$\Psi(i) = i^{-1}(\sum_{j=i+1}^{N} m_{(j)})^2 + \sum_{j=i+1}^{N} m_{(j)}^2 \quad i = 1,\ldots,N \text{ where } \quad \Psi(N) \equiv 0 .$$

Let $\tilde{\Psi}(\cdot)$ be the function that results from connecting the points $(i,\Psi(i))$ and $(i+1,\Psi(i+1))$, $i = 1,\ldots,N-1$. As we will show in Example 6.1, $\tilde{\Psi}(\cdot)$ is a decreasing convex function on $(0,N)$. Hence,

$$\inf_{p\in\bar{P}} \sum_{i=1}^{N} p(i)\tilde{\Psi}(i) = \tilde{\Psi}(n) .$$

Note that $\tilde{\Psi}(n) = \Psi(n)$ (assuming n is a positive integer $\leq N$). Therefore,

$$\inf_{p\in P_1} R(p,e_1;\lambda^{\ell}) = [E(z_1^2) - E(z_1z_2)]\Psi(n) \tag{3.4}$$

The right hand side of (3.4) is just the Bayes risk of a pair (p,e_1) where $p \in \mathcal{P}_1^*$, i.e., $\phi(\lambda^1,\ldots,\lambda^r) = \mathcal{P}_1^*$ and the proof is complete.

Note that if all the m_i's are distinct then $\mathcal{P}_1^*$ contains only one design, that is the design which puts probability one on the sample with the n largest m_i's, i.e., the good designs which we have found for $e_1(s,y)$ are essentially nonrandom in nature.

In the case when all the m_i's are equal, $e_1(s,y)$ specializes to the classical estimator namely,

$$e_2(s,y) = n^{-1}(s)N \sum_{i \in s} y_i \,.$$

Letting $\mathcal{P}_2 = \{p\colon p(s) = 0 \text{ if } s \notin S_n\}$, i.e., $\mathcal{P}_2$ is the class of designs of fixed sample size n, we see that the following corollary is an immediate consequence of Theorem 3.1.

Corollary 3.1:

For estimating the population total with squared error loss, the pair (p,e_2) where $p \in \mathcal{P}_2$ is uniformly admissible relative to $\mathcal{P}_1$.

Remark 3.1:

Meeden and Ghosh (1983) tried to show that (p,e_1) where $p \in \mathcal{P}_1^*$ is uniformly admissible relative to $\mathcal{P}_2$. In the course of the proof, they used the assumption that "neither $r_s(e_1;y)$ dominates $r_{s'}(e_1;y)$ nor vice versa for all $s, s' \in S_n$." But as we have seen in Remark 2.3, this assumption is not enough when S_n contains more than two samples. However, this uniform admissibility result is still valid. In fact, it is easy to see that it is an immediate consequence of Theorem 3.1 since $\mathcal{P}_2$ is a subset of $\mathcal{P}_1$. Similarly, the uniform admissibility of (p,e_2) relative to $\mathcal{P}_2$ is also true.

Vardeman and Meeden (1983a), have considered various estimators for estimating the population total and they have studied the admissibility of those estimators under any design and the uniform admissibility relative to the class of designs of fixed sample size n. The proof of those uniform admissibility results depend on the theorem of choosing between experiments given by Meeden and Ghosh (1983). This theorem requires either one of the following two assumptions: (i) $\forall\ s \neq s' \in S$, neither $r_s(e;y)$ dominates $r_{s'}(e;y)$ nor vice versa, or (ii) the set of priors, say $\lambda^1,\ldots,\lambda^m$ to be used is such that $\Theta = \bigcup_{j=1}^{m} \Theta(\lambda^j)$ where Θ is the (restricted) parameter space. However, as we have mentioned in the previous paragraph, assumption (i) is not enough when S contains more than two samples. On the other hand, when proving any uniform admissibility result relative to the class of designs of fixed sample size n, using the theorem given by Meeden and Ghosh (1983), only the set of samples of size n has to be considered. Accordingly, under the kind of restricted parameter space which we have talked about in Section 3.1.1, assumption (ii) cannot be satisfied. For this reason, we utilize the results given in Section 2.1.1 to give some uniform admissibility results relative to the class of designs of expected sample size less than or equal to n for those estimators. From those results, the corresponding uniform admissibility results relative to the class of designs of fixed sample size n follows immediately. First we present those estimators:

Let $m = (m_1,\ldots,m_N)$ and $X = (x_1,\ldots,x_N)$ be two vectors of known constants associated with the unknown vector $y = (y_1,\ldots,y_N)$ where $m_i \neq 0 \ \forall\ i = 1,\ldots,N$. Let $v_i = (y_i - x_i)/m_i$ for $i = 1,\ldots,N$.

Suppose that for any $i^* \notin s$, the posterior distribution is just the empirical distribution of the observed v_i. This will imply that the posterior mean of v_{i^*} is $\bar{v}_s = \frac{1}{n(s)} \sum_{i\in s} v_i$ and, hence, the posterior mean of y_{i^*} is $x_{i^*} + \bar{v}_s m_{i^*}$. This kind of argument yields the estimator

$$e_o = \sum_{i\in s} y_i + \sum_{i\notin s} x_i + \bar{v}_s \sum_{i\notin s} m_i . \tag{3.5}$$

Now, consider a situation where the Bayesian has a guess μ^* for the population mean $\bar{v}$. In such a case, μ^* can be used as a marginal mean for any unobserved v_{i^*}. This implies that $x_{i^*} + \mu^* m_{i^*}$ is a marginal mean for y_{i^*}. This kind of thinking gives the estimator

$$e_\infty = \sum_{i\in s} y_i + \sum_{i\notin s} x_i + \mu^* \sum_{i\notin s} m_i . \tag{3.6}$$

A compromise between e_o and e_∞ can be obtained by taking the posterior mean for any unobserved v_{i^*} to be a weighted average of $\bar{v}_s$ and μ^*, i.e., by taking

$$E(v_{i^*}|y(s)) = \frac{M}{M+n(s)}\,\mu^* + \frac{n(s)}{M+n(s)}\,\bar{v}_s$$

where $M \in (o,\infty)$ can be interpreted as representing how strongly the statistician believes in his prior guess μ^*. This implies,

$$E(y_{i*}|y(s)) = x_{i*} + m_{i*}[\frac{M}{M+n(s)}\mu^* + \frac{n(s)}{M+n(s)}\bar{v}_s] .$$

Hence, the compromise estimator is

$$e_M = \sum_{i\in s} y_i + \sum_{i\notin s} x_i + [\frac{M}{M+n(s)}\mu^* + \frac{n(s)}{M+n(s)}\bar{v}_s] \sum_{i\notin s} m_i . \tag{3.7}$$

Note that e_o and e_∞ are the limits of e_M as M goes to o and ∞ respectively.

We now give some special cases of the estimators e_o, e_∞ and e_M.

Case (1): Let $x_i = 0$ for all $i = 1,\ldots,N$ and let $m_1 = m_2 = \ldots = m_N$. This gives

$$e_o' = \sum_{i\in s} y_i + (N-n(s))\bar{y}_s , \tag{3.8}$$

i.e., the classical estimator where $\bar{y}_s$ is the sample mean.

If, in addition, $m_i = 1$ for all $i = 1,\ldots,N$ then

$$e_\infty' = \sum_{i\in s} y_i + (N-n(s))\mu^* \tag{3.9}$$

and

$$e_M' = \sum_{i\in s} y_i + (N-n(s))[\frac{M}{M+n(s)}\mu^* + \frac{n(s)}{M+n(s)}\bar{y}_s] . \tag{3.10}$$

Case (2): Let $x_i = 0$ for all $i = 1,\ldots,N$. This yields

$$e''_o = \sum_{i\in s} y_i + \left(\frac{1}{n(s)} \sum_{i\in s} \frac{y_i}{m_i}\right) \sum_{i\notin s} m_i \tag{3.11}$$

$$e''_\infty = \sum_{i\in s} y_i + \mu^* \sum_{i\notin s} m_i \tag{3.12}$$

and

$$e''_M = \sum_{i\in s} y_i + \left[\frac{M}{M+n(s)}\mu^* + \frac{n(s)}{M+n(s)}\bar{r}_s\right] \sum_{i\notin s} m_i \tag{3.13}$$

where $\bar{r}_s = \frac{1}{n(s)} \sum_{i\in s} \frac{y_i}{m_i}$, i.e., the sample mean of the ratios.

Case (3): Let $m_1 = m_2 = \ldots = m_N$. This implies

$$e'''_o = \sum_{i\in s} y_i + \sum_{i\notin s} x_i + (N-n(s))\bar{d}_s\,, \tag{3.14}$$

i.e., the usual difference estimator where $\bar{d}_s = \frac{1}{n(s)} \sum_{i\in s} (y_i-x_i)$.

If, in addition, $m_i = 1$ for all $i = 1,\ldots,N$ then

$$e'''_\infty = \sum_{i\in s} y_i + \sum_{i\notin s} x_i + (N-n(s))\mu^* \tag{3.15}$$

and

$$e'''_M = \sum_{i\in s} y_i + \sum_{i\notin s} x_i + (N-n(s))\left[\frac{M}{M+n(s)}\mu^* + \frac{n(s)}{M+n(s)}\bar{d}_s\right]. \tag{3.16}$$

For more discussion about the above estimators, please see Vardeman and Meeden (1983a).

Note that the estimators e_o' and e_o'' have been studied in detail in the beginning of this section. Now, we prove uniform admissibility relative to the class of designs of expected sample size less than or equal to n for e_o and e_∞.

Theorem 3.2:

For estimating the population total, $\tau(y) = \sum_{i=1}^{N} y_i$, with squared error loss, the pair (p, e_o) where $p \in P_1^*$ is uniformly admissible relative to P_1 provided $m_i > 0$ for all $i = 1,\ldots,N$.

Proof:

For any point $y^o \in \Theta$ containing distinct values $\alpha_1,\ldots,\alpha_r$ $(r \leq N)$ of the quantities $v_i = (y_i - x_i)/m_i$'s, let

$$\bar{\Theta}_v(\alpha_1,\ldots,\alpha_r) = \{y\colon v_i = \alpha_j \text{ for some } j = 1,\ldots,r; \text{ for all } i = 1,\ldots,N\}.$$

First, we show that using any sample s of size $n(s)$, where $1 \leq n(s) \leq N$, e_o is unique stepwise Bayes against some set of mutually orthogonal priors defined on $\bar{\Theta}_v(\alpha_1,\ldots,\alpha_r)$. [This will imply, by Meeden and Ghosh (1981 and 1982), that under any design, e_o is finitely admissible and, hence, is admissible.] Now, we need the following notations:

Let $\bar{\bar{\Theta}}_v(\alpha_1,\ldots,\alpha_r) = \{y\colon v_i = \alpha_j$ for some $j = 1,\ldots,r;$ for all $i = 1,\ldots,N$ and each α_j appears at least once for $j = 1,\ldots,r\}$. If $y \in \bar{\bar{\Theta}}_v(\alpha_1,\ldots,\alpha_r)$ we say that y is of order r for $\alpha_1,\ldots,\alpha_r$. Similarly, if $y(s)$ is a sample point with $r \le n(s)$ we say that $y(s)$ is of order r for $\alpha_1,\ldots,\alpha_r$ if each v_i equals one of the r values $\alpha_1,\ldots,\alpha_r$ and if for each value α_j, there exists at least one i_ℓ for which $v_{i_\ell} = \alpha_j$. If $y \in \bar{\bar{\Theta}}_v(\alpha_1,\ldots,\alpha_r)$ let $w_v(y;j)$ be the number of v_i's which are equal to α_j. Note that for each j, $w_v(y;j) \ge 1$ and $\sum_{j=1}^{r} w_v(y;j) = N$. If $y(s)$ is a sample point of order r for $\alpha_1,\ldots,\alpha_r$ let $w_v(s;j)$ be the number of observed v_i's which are equal to α_j. It is clear that $w_v(s;j) \ge 1$ and $\sum_{j=1}^{r} w_v(s;j) = n(s)$.

Let s be a typical sample of size $n(s)$ where $1 \le n(s) \le N$. Now, we define a set of mutually orthogonal priors against which e_o is unique stepwise Bayes.

Let λ^1 be a prior that assigns mass $\frac{1}{r}$ to each point in the set $\bigcup_{j=1}^{r} \bar{\bar{\Theta}}_v(\alpha_j)$. The sample points that have positive marginal probability under this prior are those of order one for some α_j. For any such point, say $y(s)$, and any $i^*\notin s$ we have

$$p[v_{i^*} = \alpha_j | y(s)] = 1$$

and, hence,

$$E(y_{i^*} | y(s)) = m_{i^*}\alpha_j + x_{i^*}\,.$$

Therefore, the Bayes estimate at $y(s)$ is

$$E[\tau|y(s)] = \sum_{i\in s} y_i + \sum_{i\notin s} E[y_i|y(s)]$$

$$= \sum_{i\in s} y_i + \sum_{i\notin s} x_i + \alpha_j \sum_{i\notin s} m_i$$

which is just e_o at $y(s)$.

Now, define λ^2 on the set $\bigcup_{\{\ell<j\}} \bar{\bar{\Theta}}_v(\alpha_\ell,\alpha_j)$ as follows:

$$\lambda^2(y) \propto \int_o^1 \nu^{w_v(y;\ell)-1} (1-\nu)^{w_v(y;j)-1} d\nu$$

$$= \frac{\Gamma[w_v(y;\ell)]\ \Gamma[w_v(y;j)]}{\Gamma[N]} .$$

The data points that have positive marginal probability under λ^2, but not under λ^1, are those of order two for some α_ℓ and α_j, $\ell < j$. For any such point, $y(s)$, the marginal probability is

$$\lambda^2(y(s)) \propto \Gamma[w_v(s;\ell)]\ \Gamma[w_v(s;j)]/\Gamma[n(s)] .$$

Hence, for any $i^* \notin s$, we get

$$p[v_{i*} = \alpha_\ell|y(s)] = \lambda^2[y(s) \text{ and } v_{i*} = \alpha_\ell]/\lambda^2(y(s))$$

$$= w_v(s;\ell)/n(s) .$$

Therefore,

$$E[v_{i*}|y(s)] = [\alpha_\ell w_v(s;\ell) + \alpha_j w_v(s;j)]/n(s)$$

and, hence, the Bayes estimate at $y(s)$ is

$$E[\tau|y(s)] = \sum_{i\in s} y_i + \sum_{i\notin s} E(y_i|y(s))$$

$$= \sum_{i\in s} y_i + \sum_{i\notin s} \{x_i + m_i[\alpha_\ell w_v(s;\ell) + \alpha_j w_v(s;j)]/n(s)\}$$

which is just e_o at $y(s)$.

Define the third prior λ^3 on the set $\bigcup_{\{i<j<k\}} \bar{\bar{\Theta}}_v(\alpha_i,\alpha_j,\alpha_k)$ as follows:

$$\lambda^3(y) \propto \int_o^1\int_o^1 \nu_1^{w_v(y;i)-1}\, \nu_2^{w_v(y;j)-1}\, (1-\nu_1-\nu_2)^{w_v(y;k)-1}\, d\nu_1 d\nu_2 .$$

The sample points which have positive marginal probability under λ^3 but not under λ^1 or λ^2 are those of order three for some α_i, α_j and α_k. As before, for any such point, it can be shown that the Bayes estimate under λ^3 is just e_o at this sample point.

The rest of the priors $\lambda^4, \lambda^5, \ldots, \lambda^r$ can be defined in an analogue way. In general, λ^ℓ will be defined on the set $\bigcup_{\{i_1<i_2<\ldots<i_\ell\}} \bar{\bar{\Theta}}_v(\alpha_{i_1},\ldots,\alpha_{i_\ell})$ as follows:

$$\lambda^\ell(y) \propto \int_o^1 \ldots \int_o^1 \prod_{j=1}^{\ell} \nu_j^{w_v(y;i_j)-1} \prod_{j=1}^{\ell} d\nu_j$$

and the data points that have positive marginal probability under λ^ℓ

but not under λ^k for any $k < \ell$ are those of order ℓ for some $\alpha_{i_1},\ldots,\alpha_{i_\ell}$. For any such point, the posterior probability that an unobserved v_{i*} takes on the value α_{i_j} is $w_v(s;i_j)/n(s)$ and, hence, e_o can be identified as the Bayes estimate against λ^ℓ at this data point.

Hence, e_o is unique stepwise Bayes against $\lambda^1,\ldots,\lambda^r$. [This implies, by Meeden and Ghosh (1981 and 1982), that under any design, e_o is finitely admissible and, hence, is admissible.]

Next, we compute the Bayes risks against $\lambda^1,\ldots,\lambda^r$ of a pair (p,e_o) where $p \in P_1$. For a typical sample s of size $n(s)$, the risk function is

$$r_s(e_o;y) = [\sum_{i=1}^{N} y_i - \sum_{i\in s} y_i - \sum_{i\notin s} x_i - \bar{v}_s \sum_{i\notin s} m_i]^2$$

$$= [\sum_{i\notin s} (y_i - x_i - \bar{v}_s m_i)]^2$$

$$= [\sum_{i\notin s} (m_i v_i) - (\frac{1}{n(s)} \sum_{i\in s} v_i)(\sum_{i\notin s} m_i)]^2$$

$$= [\sum_{i=1}^{N} a_{i,s} v_i]^2$$

where $a_{i,s} = -[n(s)]^{-1} \sum_{i\notin s} m_i$ for all $i\in s$ and $a_{i,s} = m_i$ for all $i\notin s$. Hence,

$$r_s(e_o;y) = \sum_{i=1}^{N} a_{i,s}^2 v_i^2 + \sum_{\substack{i=1\\ i\neq j}}^{N} \sum_{j=1}^{N} a_{i,s} a_{j,s} v_i v_j$$

Note that the Bayes risk of e_o under λ^1 is zero. Now, under any of the priors $\lambda^2,\ldots,\lambda^r$, the v_i's are finitely exchangeable and consequently the marginal prior distributions of the v_i's are all the same and the joint marginal prior of any pair (v_i,v_j) is the same as that of (v_1,v_2). Hence, under λ^ℓ, the Bayes risk of e_o based on a sample s of size $n(s)$ is

$$R_s(e_o;\lambda^\ell) = \sum_{i=1}^{N} a_{i,s}^2 E(v_1^2) + \sum_{i\neq j}\sum a_{i,s}a_{j,s}E(v_1v_2)$$

where the expectations are taken with respect to the marginal priors under λ^ℓ. Note that $\sum_{i=1}^{N} a_{i,s} = 0$. Therefore,

$$R_s(e_o;\lambda^\ell) = [E(v_1^2) - E(v_1v_2)] \sum_{i=1}^{N} a_{i,s}^2 .$$

Now, the Bayes risk of a pair (p,e_o) with $p \in \mathcal{P}_1$ is

$$R(p,e_o;\lambda^\ell) = \sum_{s\in S} p(s)R_s(e_o;\lambda^\ell)$$

$$= [E(v_1^2) - E(v_1v_2)] \sum_{s\in S} p(s) \sum_{i=1}^{N} a_{i,s}^2$$

$$= [E(v_1^2) - E(v_1v_2)] \sum_{s\in S} p(s)[\frac{1}{n(s)} (\sum_{i\notin s} m_i)^2 + \sum_{i\notin s} m_i^2].$$

By the Schwarz inequality, we have $E(v_1^2) - E(v_1v_2) \geq 0$. Hence,

$$\inf_{p\in P_1} R(p,e_o;\lambda^{\ell}) = [E(v_1^2)-E(v_1v_2)] \inf_{p\in P_1} \sum_{s\in S} p(s)[\frac{1}{n(s)}(\sum_{i\notin s} m_i)^2 + \sum_{i\notin s} m_i^2]. \tag{3.17}$$

Note that equation (3.17) is exactly equation (3.3) with v_1 and v_2 replacing z_1 and z_2. Hence, the result follows using the same steps given after equation (3.3), and the proof is complete.

Now, if all m_i's are equal then P_1^* is just P_2 and, hence, the following corollary is immediate.

Corollary 3.2:

For estimating the population total with squared error loss we have

(i) (p,e_o') where $p \in P_2$ is uniformly admissible relative to P_1.

(ii) (p,e_o'') where $p \in P_1^*$ is uniformly admissible relative to P_1 provided $m_i > 0$ for all $i = 1,\ldots,N$.

(iii) (p,e_o''') where $p \in P_2$ is uniformly admissible relative to P_1.

Note that the results of (i) and (ii) were given previously in Theorem 3.1 and Corollary 3.1.

Now, the following theorem identifies some designs so that when used with e_∞ then the resulting pair is uniformly admissible relative to the class of designs of expected sample size less than or equal to n.

Theorem 3.3:

For estimating the population total, $\tau(y) = \sum_{i=1}^{N} y_i$, with squared error loss, the pair (p, e_∞) where $p \in \mathcal{P}_1^*$ is uniformly admissible relative to $\mathcal{P}_1$ provided $m_i > 0$ for all $i = 1,\ldots,N$.

Proof:

For any point $y^o \in \Theta$ containing distinct values $a_1,\ldots,a_{r^*}$ $(r^* \leq N)$ for the v_i's there exists a set of real numbers $A = \{\alpha_1,\ldots,\alpha_r\}$ $(r < \infty)$ where $\{\alpha_1,\ldots,\alpha_r\} \supset \{a_1,\ldots,a_{r^*}\}$ and a probability distribution $\pi = (\pi_1,\ldots,\pi_r)$ on $\{\alpha_1,\ldots,\alpha_r\}$ such that $\pi_i > 0$ for all $i = 1,\ldots,r$ and $\sum_{i=1}^{r} \alpha_i \pi_i = \mu^*$. Let $A_1 = \{y: v_j \in A \text{ for } j = 1,\ldots,N\}$. Taking A_1^N to be the restricted parameter space, we first show that for any sample s of size $n(s)$ where $1 \leq n(s) \leq N$, e_∞ is unique Bayes against some prior distribution. [This will imply, by Meeden and Ghosh (1982), that under any design e_∞ is finitely admissible and, hence, is admissible.]

For any $y \in A_1^N$, let $w_v(y;\ell)$ be the number of v_i's that are equal to α_ℓ and define the prior distribution λ on A_1^N as follows:

$$\lambda(y) = \prod_{\ell=1}^{r} \pi_\ell^{w_v(y;\ell)}$$

For a sample point $y(s)$, let $w_v(s;\ell)$ be the number of v_i's $i \in s$ that are equal to α_ℓ. Hence, the marginal probability for $y(s)$ is

$$\lambda(y(s)) = \prod_{\ell=1}^{r} \pi_\ell^{w_v(s;\ell)} .$$

Now, for any $i^* \notin s$ and any $k = 1,\ldots,r$ we have

$$p(v_{i^*} = \alpha_k | y(s)) = \lambda[y(s) \text{ and } v_{i^*} = \alpha_k]/\lambda(y(s)) = \pi_k .$$

Hence,

$$E[y_{i^*}|y(s)] = x_{i^*} + m_{i^*}\,\mu^*$$

and, consequently, the Bayes estimate at $y(s)$ is

$$E[\tau|y(s)] = \sum_{i\in s} y_i + \sum_{i\notin s} E[y_i|y(s)]$$

$$= \sum_{i\in s} y_i + \sum_{i\notin s} x_i + \mu^* \sum_{i\notin s} m_i$$

which is just e_∞, i.e., e_∞ is unique Bayes against λ when the parameter space is taken to be A^N. [This implies that under any design, e_∞ is finitely admissible and, hence, is admissible.]

Next, we compute the Bayes risk against λ of a pair (p,e_∞) with $p \in \mathcal{P}_1$. For a typical sample s of size $n(s)$, the risk function is

$$r_s(e_\infty;y) = [\sum_{i=1}^{N} y_i - \sum_{i\in s} y_i - \sum_{i\notin s} x_i - \mu^* \sum_{i\notin s} m_i]^2$$

$$= [\sum_{i\notin s} (y_i - x_i - \mu^* m_i)]^2$$

$$= [\sum_{i \notin s} m_i (v_i - \mu^*)]^2$$

$$= \sum_{i \notin s} m_i^2 (v_i - \mu^*)^2 + \sum_{i \notin s} \sum_{\substack{j \notin s \\ i \neq j}} m_i m_j (v_i - \mu^*)(v_j - \mu^*) .$$

Note that under the prior λ, the v_i's are independent and identically distributed and the expected value of any of the v_i's is μ^*. Hence, the Bayes risk against λ of e_∞ based on a sample s is

$$R_s(e_\infty;\lambda) = E(v_1 - \mu^*)^2 \sum_{i \notin s} m_i^2$$

where the expectation is taken with respect to the marginal prior under λ.

Now, the Bayes risk under λ of a pair (p, e_∞) with $p \in \mathcal{P}_1$ is

$$R(p, e_\infty;\lambda) = \sum_{s \in S} p(s) R_s(e_\infty;\lambda)$$

$$= E(v_1 - \mu^*)^2 \sum_{s \in S} p(s) \sum_{i \notin s} m_i^2 .$$

Consequently,

$$\inf_{p \in \mathcal{P}_1} R(p, e_\infty;\lambda) = E(v_1 - \mu^*)^2 \inf_{p \in \mathcal{P}_1} \sum_{s \in S} p(s) \sum_{i \notin s} m_i^2 . \tag{3.18}$$

By ordering the m_i's such that $m_{(1)} \geq m_{(2)} \geq \cdots \geq m_{(N)}$ and letting $\bar{S}_{n(s)} = \{s\colon s \in S_{n(s)}$ and $\sum_{i \in s} m_i = \sum_{j=1}^{n(s)} m_{(j)}\}$ where $S_{n(s)}$ is the set of samples of size $n(s)$, $\bar{S} = \bigcup_{n(s)=1}^{N} \bar{S}_{n(s)}$ and $\bar{P} = \{p\colon p \in P_1$ and $p(s) = 0$ if $s \notin \bar{S}\}$ we note that the set of designs that gives the infimum of (3.18) is a subset of $\bar{P}$. Therefore,

$$\inf_{p \in P_1} R(p, e_\infty; \lambda) = E(v_1 - \mu^*)^2 \inf_{p \in \bar{P}} \sum_{s \in \bar{S}} p(s) \sum_{i \notin s} m_i^2$$

$$= E(v_1 - \mu^*)^2 \inf_{p \in \bar{P}} \sum_{i=1}^{N} p(i) \Psi_2(i)$$

where $p(i)$ is the probability, under the design p, of selecting the sample of size i that has the largest m_i's and $\Psi_2(i)$ is

$$\Psi_2(i) = \sum_{j=i+1}^{N} m_{(j)}^2 \qquad i = 1, \ldots, N \text{ where } \Psi_2(N) \equiv 0 .$$

Let $\tilde{\Psi}_2(\cdot)$ be the function that results from connecting the points $(i, \Psi_2(i))$ and $(i+1, \Psi_2(i+1))$, $i = 1, \ldots, N-1$. As we will show in example (6.2) in the Appendix, $\tilde{\Psi}_2(\cdot)$ is convex and strictly decreasing on $(0,N)$. Hence,

$$\inf_{p \in \bar{P}} \sum_{i=1}^{N} p(i) \tilde{\Psi}_2(i) = \tilde{\Psi}_2(n) .$$

Note that $\tilde{\Psi}_2(n) = \Psi_2(n)$ (assuming n is a positive interger $\leq N$). Therefore,

$$\inf_{p \in P_1} R(p, e_\infty; \lambda) = E(v_1 - \mu^*)^2 \Psi_2(n) . \tag{3.19}$$

The right hand side of (3.19) is just the Bayes risk of a pair (p,e_∞) where $p \in \mathcal{P}_1^*$. Hence, (p,e_∞) where $p \in \mathcal{P}_1^*$ is finitely uniformly admissible relative to $\mathcal{P}_1$ and, hence, by Meeden and Ghosh (1982) it is uniformly admissible relative to $\mathcal{P}_1$ and the proof is complete.

Now, the following corollary gives the above result in the special cases of e_∞. (Recall that if all m_i's are equal then $\mathcal{P}_1^*$ is just $\mathcal{P}_2$.)

<u>Corollary 3.3</u>:

For estimating the population total with squared error loss, we have

(i) (p,e_∞') where $p \in \mathcal{P}_2$ is uniformly admissible relative to $\mathcal{P}_1$.

(ii) (p,e_∞'') where $p \in \mathcal{P}_1^*$ is uniformly admissible relative to $\mathcal{P}_1$ provided that $m_i > 0$ for all $i = 1,\ldots,N$.

(iii) (p,e_∞''') where $p \in \mathcal{P}_2$ is uniformly admissible relative to $\mathcal{P}_1$.

<u>Remark 3.2</u>:

Note that the uniform admissibility results given in Theorems 3.2 and 3.3 (and, consequently, their special cases given in Corollaries 3.2 and 3.3) are true if the class of designs $\mathcal{P}_1$ is replaced by any subset of $\mathcal{P}_1$ that contains $\mathcal{P}_2$. In particular, those uniform admissibility results are true if we replace $\mathcal{P}_1$ by $\mathcal{P}_2$, the class of designs of fixed sample size n.

Vardeman and Meeden (1983a) conjectured that both (p,e_∞) and (p,e''_∞) where $p \in \mathcal{P}_1^*$ are uniformly admissible relative to the class of designs of fixed sample size n. This conjecture is, in fact, supported by Theorem 3.3 and part (ii) of Corollary 3.3.

We have been, so far, unsuccessful to give any uniform admissibility results concerning e_M or any of its special cases.

The particular estimators of τ presented in this section have the virtue that they are relatively simple and intuitively reasonable ways to make use of the kinds of prior information that can, in some cases, be available in a sampling problem.

From the line of argument, the estimators considered in this section have been established, we see that the possibility of introducing other estimators of this type is endless. For instance, Vardeman and Meeden (1983a) have considered this type of estimators in the more general case of having N known 1-1 functions from $\mathbb{R}$ onto $\mathbb{R}$, say $\zeta_1, \zeta_2, \ldots, \zeta_N$, and they have established the admissibility of such estimators under any design. However, nothing can be said about uniform admissibility of those estimators since this issue depends on the form of the functions $\zeta_1, \zeta_2, \ldots, \zeta_N$ to be considered. (Note that in this section, the special form $\zeta_i(y_i) = \frac{y_i - x_i}{m_i}$ is considered.)

Remark 3.3:

Note that the previous uniform admissibility results concerning e_o and e_∞ (and their special cases e''_o and e''_∞) have been given in the case when all the m_i's are positive. However, some uniform admissibility results can be obtained as well for those estimators when all the m_i's need not be positive.

If $m_i < 0$ for all $i = 1,\ldots,N$, then by ordering the m_i's such that $m_{(1)} \leq m_{(2)} \leq \cdots \leq m_{(N)}$ and following the proof of Theorems 3.2 and 3.3, we see that the pairs (p,e_o), (p,e_o''), (p,e_∞) and (p,e_∞'') where $p \in \mathcal{P}_1'$ are uniformly admissible relative to $\mathcal{P}_1$ where

$$\mathcal{P}_1' = \{p\colon p \in \mathcal{P}_1 \text{ and } p(s) = 0 \text{ if } s \notin S'\} \text{ and}$$

$$S' = \{s\colon s \in S_n \text{ and } \sum_{i\in s} m_i = \min_{s'\in S_n} \sum_{i\in s'} m_i\}.$$

If some of the m_i's are positive and some are negative, uniform admissibility results concerning e_∞ and e_∞'' can be obtained by ordering the m_i's such that $m_{(1)}^2 \geq m_{(2)}^2 \geq \cdots \geq m_{(N)}^2$ and following the proof of Theorem 3.3. In this case, we see that the pairs (p,e_∞) and (p,e_∞'') where $p \in \mathcal{P}_1''$ are uniformly admissible relative to $\mathcal{P}_1$ where

$$\mathcal{P}_1'' = \{p\colon p \in \mathcal{P}_1 \text{ and } p(s) = 0 \text{ if } s \notin S''\} \text{ and}$$

$$S'' = \{s\colon s \in S_n \text{ and } \sum_{i\in s} m_i^2 = \max_{s'\in S_n} \sum_{i\in s'} m_i^2\}.$$

3.1.3. Admissibility of a U-statistic when estimating the population counterpart

For any $k \leq N$ and any symmetric function $\xi(.,\ldots,.)$ let

$$U_p(y_1,\ldots,y_N) = [1/\binom{N}{k}] \sum_{\beta \in B} \xi(y_{\beta_1},\ldots,y_{\beta_k}) \tag{3.20}$$

where $B = \{\beta | \beta = (\beta_1,\ldots,\beta_k)$ is one of the $\binom{N}{k}$ unordered subsets of k integers chosen without replacement from the set $\{1,\ldots,N\}\}$.

$U_p(y_1,\ldots,y_N)$ is a class of parametric functions of the population whose sample counterpart, called "U-statistic," is defined (for a given sample s of size $n(s) \geq k$) as follows:

$$U_s(y_{i_1},\ldots,y_{i_{n(s)}}) = [1/\binom{n(s)}{k}] \sum_{\beta^k \in B^k} \xi(y_{\beta_1^k},\ldots,y_{\beta_k^k}) \tag{3.21}$$

where $B^k = \{\beta^k | \beta^k = (\beta_1^k,\ldots,\beta_k^k)$ is one of the $\binom{n(s)}{k}$ unordered subsets of k integers chosen without replacement from the set $\{1,\ldots,n(s)\}\}$.

Note that a U-statistic is symmetric in its arguments. Moreover, it has some nice properties when choosing the sample randomly from a population with some distributional assumptions (please see Randles and Wolfe (1979)).

Our interest in this section is to construct an admissible estimator for $U_p(y_1,\ldots,y_N)$. The following theorem provides this estimator which is, in fact, a proper multiple of $U_s(y_{i_1},\ldots,y_{i_{n(s)}})$.

Theorem 3.4:

Under squared error loss and any design such that $n(s) \geq k$ with probability one, an admissible estimator of

$$U_p(y_1,\ldots,y_N) = [\binom{N}{k}]^{-1} \sum_{\beta \in B} \xi(y_{\beta_1},\ldots,y_{\beta_k}),$$

where $\xi(.,\ldots,) = 0$ if two or more of its coordinates are equal, is given by

$$U^*(y_{i_1},\ldots,y_{i_{n(s)}}) = [\binom{n(s)}{k}/\binom{N}{k}]\{1 + \sum_{j=1}^{k} \binom{N-n(s)}{j}$$

$$\frac{[k(k-1)\ldots(k-j+1)]}{[n(s)(n(s)+1)\ldots(n(s)+j-1)]}\} U_s(y_{i_1},\ldots,y_{i_{n(s)}}).$$

Proof:

For any point $y^o \in \Theta$ containing distinct values $\alpha_1,\ldots,\alpha_r$ $(r \leq N)$ let $\bar{\Theta}(\alpha_1,\ldots,\alpha_r) = \{y\colon y_i = \alpha_j$ for some $j = 1,\ldots,r;$ for all $i = 1,\ldots,N\}$.

Taking the parameter space to be $\bar{\Theta}(\alpha_1,\ldots,\alpha_r)$ we now show that, under any design with $n(s) \geq k$ and squared error loss, $U^*(y_{i_1},\ldots,y_{i_{n(s)}})$ is unique stepwise Bayes against some set of mutually orthogonal priors. This will imply, by Meeden and Ghosh (1981 and 1982), that under any design, $U^*(y_{i_1},\ldots,y_{i_{n(s)}})$ is finitely admissible and, hence, is admissible.

First, we need the following notations:

Let $\bar{\bar{\Theta}}(\alpha_1,\ldots,\alpha_r) = \{y\colon y_i = \alpha_j$ for some $j = 1,\ldots,r;$ for all

$i = 1,\ldots,N$ and each α_j appears at least once for $j = 1,\ldots,r\}$. If $y \in \bar{\bar{\Theta}}(\alpha_1,\ldots,\alpha_r)$ we say that y is of order r for $\alpha_1,\ldots,\alpha_r$. Similarly, if $y(s)$ is a sample point with $r \leq n(s)$ we say that $y(s)$ is of order r for $\alpha_1,\ldots,\alpha_r$ if each y_i equals one of the r values $\alpha_1,\ldots,\alpha_r$ and if for each value α_j, there exists at least one i_ℓ for which $y_{i_\ell} = \alpha_j$. If $y \in \bar{\bar{\Theta}}(\alpha_1,\ldots,\alpha_r)$, let $w_y(j)$ be the number of y_i's which are equal to α_j. Note that for each j, $w_y(j) \geq 1$ and $\sum_{j=1}^{r} w_y(j) = N$. If $y(s)$ is a sample point of order r for $\alpha_1,\ldots,\alpha_r$ let $w_y(j;s)$ be the number of observed y_i's $(i \in s)$ which are equal to α_j. It is clear that $w_y(j;s) \geq 1$ and $\sum_{j=1}^{r} w_y(j;s) = n(s)$.

Let s be some fixed sample of size $n(s) \geq k$. Note that $U_p(y_1,\ldots,y_N)$ can be rewritten as

$$U_p(y_1,\ldots,y_N) = [1/\binom{N}{k}] \sum_{i=0}^{k} \sum_{\beta^i \in B^i} \xi(y_{\beta_1^i},\ldots,y_{\beta_k^i}) \tag{3.22}$$

where $B^i = \{\beta^i | \beta^i = (\beta_1^i,\ldots,\beta_k^i)$ is one of the $\binom{n(s)}{i}\binom{N-n(s)}{k-i}$ unordered subsets of k integers chosen without replacement from the set $\{1,\ldots,N\}$ where i of them chosen from the set $\{i_1,\ldots,i_{n(s)}\}$ and $k-i$ chosen from the set $\{1,\ldots,N\} \cap \{i_1,\ldots,i_{n(s)}\}^c\}$.

Recall that under squared error loss, the Bayes estimate at an observed sample, $y(s)$, against some prior is just the posterior mean.

We now present a set of mutually orthogonal priors against which $U^*(y_{i_1},\dots,y_{i_{n(s)}})$ is unique stepwise Bayes.

The first prior λ^1 assigns mass $\frac{1}{r}$ to each point in the set $\bigcup_{j=1}^{r} \bar{\bar{\Theta}}(\alpha_j)$. The data points that are consistent under this prior are those where all the observed values are equal. In this case, the assumption on the function $\xi(.,\dots,.)$ implies that

$$E[U_p(y_1,\dots,y_N)|y(s)] = 0 = U^*(y_{i_1},\dots,y_{i_{n(s)}}).$$

It is obvious that this will also be the case when defining a set of mutually orthogonal priors $\lambda^2,\dots,\lambda^{k-1}$ respectively on the sets $\bigcup_{\{j_1<j_2\}} \bar{\bar{\Theta}}(\alpha_{j_1},\alpha_{j_2})$, $\bigcup_{\{j_1<j_2<j_3\}} \bar{\bar{\Theta}}(\alpha_{j_1},\alpha_{j_2},\alpha_{j_3})$, ..., $\bigcup_{\{j_1<j_2<\dots<j_{k-1}\}} \bar{\bar{\Theta}}(\alpha_{j_1},\dots,\alpha_{j_{k-1}})$.

Next, the prior λ^k is defined on the set $\bar{\bar{\Theta}}(\alpha_1,\dots,\alpha_k)$ as follows:

$$\lambda^k(y) \propto \int_0^1 \int_0^1 \dots \int_0^1 \prod_{j=1}^{k} \nu_j^{w_y(j)-1} \prod_{j=1}^{k} d\nu_j = \prod_{j=1}^{k} \Gamma(w_y(j))/\Gamma(N).$$

The data points that are consistent under this prior are those of order less than or equal to k. However, the data points of order less than k have been taken care of. Now, for a data point of order k for $\alpha_1,\dots,\alpha_k$, the marginal probability of $y(s)$ is given by

$$\lambda^k(y(s)) \propto [\prod_{j=1}^{k} \Gamma(w_y(j;s))]/\Gamma(n(s))$$

Let $D^{k-i} = (D_1^{k-i}, D_2^{k-i}, \ldots, D_{k-i}^{k-i})$ be a subset of the $\binom{N-n(s)}{k-i}$ unordered subsets of $k-i$ integers chosen without replacement from the set $\{i_1,\ldots,i_{n(s)}\}^c \cap \{1,\ldots,N\}$. Hence, for any $j_1 \neq j_2 \neq \ldots \neq j_{k-i}$ we have

$$p(y_{D_1^{k-i}} = \alpha_{j_1}, y_{D_2^{k-i}} = \alpha_{j_2},\ldots,y_{D_{k-i}^{k-i}} = \alpha_{j_{k-i}} | y(s))$$

$$= \lambda^k(y(s) \text{ and } y_{D_1^{k-i}} = \alpha_{j_1}, y_{D_2^{k-i}} = \alpha_{j_2},\ldots,y_{D_{k-i}^{k-i}} = \alpha_{j_{k-i}})/\lambda^k(y(s))$$

$$= [\frac{\prod_{\ell=1}^{k-i} \Gamma(w_y(j_\ell;s)+1) \prod_{\ell=k-i+1}^{k} \Gamma(w_y(j_\ell;s))}{\Gamma(n(s)+k-i)}]/[\frac{\prod_{\ell=1}^{k} \Gamma(w_y(j_\ell;s))}{\Gamma(n(s))}]$$

$$= \frac{\prod_{\ell=1}^{k-i} w_y(j_\ell;s)}{n(s)(n(s)+1) \ldots (n(s)+k-i-1)}$$

Hence, the Bayes estimate under λ^k is

$$E[U_p(y_1,\ldots,y_N)|y(s)] = [1/\binom{N}{k}] \; E[\sum_{i=0}^{k} \sum_{\beta^i \in B^i} \xi(y_{\beta_1^i},\ldots,y_{\beta_k^i})|y(s)]$$

$$= [1/\binom{N}{k}][\sum_{\beta^k \in B^k} \xi(y_{\beta_1^k},\ldots,y_{\beta_k^k})$$

$$+ E\{\sum_{i=0}^{k-1} \sum_{\beta^i \in B^i} \xi(y_{\beta_1^i},\ldots,y_{\beta_k^i})|y(s)\}]. \qquad (3.23)$$

Without loss of generality, assume that the first i coordinates of ξ in the second term of (3.23) are the observed values. Therefore,

$$E[\sum_{i=0}^{k-1} \sum_{\beta^i \in B^i} \xi(y_{\beta_1^i},\ldots,y_{\beta_k^i})|y(s)]$$

$$= E[\sum_{i=0}^{k-1} \sum_{D^{k-i}\in \mathcal{D}^{k-i}} \sum_{C^i \in \mathcal{C}^i} \xi(y_{C_1^i},\ldots,y_{C_i^i}, y_{D_1^{k-i}},\ldots,y_{D_{k-i}^{k-i}})|y(s)]$$

(3.24)

where $\mathcal{C}^i = \{C^i | C^i = (C_1^i,\ldots,C_i^i)$ is one of the $\binom{n(s)}{i}$ unordered subsets of i integers chosen without replacement from the set $\{i_1,\ldots,i_{n(s)}\}\}$, and $\mathcal{D}^{k-i} = \{D^{k-i} | D^{k-i} = (D_1^{k-i},\ldots,D_{k-i}^{k-i})$ is one of the $\binom{N-n(s)}{k-i}$ unordered subsets of $k-i$ integers chosen without replacement from the set $\{i_1,\ldots,i_{n(s)}\}^c \cap \{1,\ldots,N\}\}$. Now, (3.24) can be rewritten as

$$\sum_{i=0}^{k-1} \sum_{D^{k-i}\in \mathcal{D}^{k-i}} \sum_{C^i \in \mathcal{C}^i} \{E[\xi(y_{C_1^i},\ldots,y_{C_i^i}, y_{D_1^{k-i}},\ldots,y_{D_{k-i}^{k-i}})|y(s)]\}$$

$$= \sum_{i=0}^{k-1} \sum_{D^{k-i}\in \mathcal{D}^{k-i}} \sum_{C^i \in \mathcal{C}^i} [\sum_{\substack{j_1=1 \\ j_1 \neq j_2 \neq \cdots \neq j_{k-i}}}^{k} \sum_{j_2=1}^{k} \cdots \sum_{j_{k-i}=1}^{k}$$

$$\xi(y_{c_1^i},\ldots,y_{c_i^i},\alpha_{j_1},\alpha_{j_2},\ldots,\alpha_{j_{k-i}})\frac{\prod_{\ell=1}^{k-i} w_y(j_\ell;s)}{n(s)(n(s)+1)\ldots(n(s)+k-i-1)}]$$

$$=\sum_{i=0}^{k-1}\binom{N-n(s)}{k-i}\sum_{c^i\in C^i}\{\sum_{c_{i+1}\in s}\ \sum_{c_{i+2}\in s}\ldots\sum_{c_k\in s}$$

$$\xi(y_{c_1^i},\ldots,y_{c_i^i},y_{c_{i+1}},y_{c_{i+2}},\ldots,y_{c_k})/[n(s)(n(s)+1)\ldots(n(s)+k-i-1)]\}$$

$$=\sum_{i=0}^{k-1}\binom{N-n(s)}{k-i}\frac{k!}{i!}\sum\sum\ldots\sum_{c_1^i<c_2^i<\ldots<c_i^i<c_{i+1}<c_{i+2}<\ldots<c_k}\sum\sum\ldots\sum$$

$$\xi(y_{c_1^i},\ldots,y_{c_i^i},y_{c_{i+1}},\ldots,y_{c_k})/[n(s)(n(s)+1)\ldots(n(s)+k-i-1)]$$

$$=\sum_{i=0}^{k-1}\binom{N-n(s)}{k-i}\frac{k!}{i!\{n(s)(n(s)+1)\ldots(n(s)+k-i-1)\}}\binom{n(s)}{k}U_s(y_{i_1},\ldots,y_{i_{n(s)}})$$

(3.25)

Substituting with (3.25) in (3.23), we get

$$E[U_p(y_1,\ldots,y_N)|y(s)]=\frac{\binom{n(s)}{k}}{\binom{N}{k}}\{1+\sum_{i=0}^{k-1}\frac{\binom{N-n(s)}{k-i}k!}{[n(s)(n(s)+1)\ldots(n(s)+k-i-1)]i!}\}$$

$$U_s(y_{i_1},\ldots,y_{i_{n(s)}}).$$

Letting $j = k-i$, we get

$$E[U_p(y_1,\ldots,y_N)\mid y(s)] = \frac{\binom{n(s)}{k}}{\binom{N}{k}}\left\{1 + \sum_{j=1}^{k} \frac{\binom{N-n(s)}{j}\, k!}{[n(s)(n(s)+1)\ldots(n(s)+j-1)](k-j)!}\right\}$$

$$U_s(y_{i_1},\ldots,y_{i_{n(s)}})$$

$$= \frac{\binom{n(s)}{k}}{\binom{N}{k}}\left\{1 + \sum_{j=1}^{k} \binom{N-n(s)}{j} \frac{[k(k-1)\ldots(k-j+1)]}{[n(s)(n(s)+1)\ldots(n(s)+j-1)]}\right\}$$

$$U_s(y_{i_1},\ldots,y_{i_{n(s)}})$$

which is $U^*(y_{i_1},\ldots,y_{i_{n(s)}})$.

Note that this will also be the case when defining any prior of the type of λ^k on any set $\bar{\bar{\Theta}}(\alpha_{i_1},\ldots,\alpha_{i_k})$. In fact, it would have been better if we defined λ^k on the set $\bigcup_{\{i_1<i_2<\ldots<i_k\}} \bar{\bar{\Theta}}(\alpha_{i_1},\alpha_{i_2},\ldots,\alpha_{i_k})$ and the proof would have been exactly the same as above. However, to avoid the complexity of notations, we defined λ^k just on $\bar{\bar{\Theta}}(\alpha_1,\ldots,\alpha_k)$.

Assume that λ^k was defined on $\bigcup_{\{i_1<i_2<\ldots<i_k\}} \bar{\bar{\Theta}}(\alpha_{i_1},\ldots,\alpha_{i_k})$ (otherwise there will be $\binom{r}{k}$ priors defined on the sets that are subsets of $\bigcup_{\{i_1<i_2<\ldots<i_k\}} \bar{\bar{\Theta}}(\alpha_{i_1},\ldots,\alpha_{i_k})$) and define the next prior λ^{k+1} (or $\lambda^{k+\binom{r}{k}}$) on $\bigcup_{\{i_1<i_2<\ldots<i_{k+1}\}} \bar{\bar{\Theta}}(\alpha_{i_1},\ldots,\alpha_{i_{k+1}})$ (or on $\bar{\bar{\Theta}}(\alpha_1,\ldots,\alpha_{k+1})$) as follows:

$$\lambda^{k+1}(y) \propto \int_0^1 \int_0^1 \cdots \int_0^1 \prod_{\ell=1}^{k+1} \nu_{j_\ell}^{w_y(j_\ell)-1} \prod_{\ell=1}^{k+1} d\nu_{j_\ell} = \prod_{\ell=1}^{k+1} \Gamma(w_y(j_\ell))/\Gamma(N).$$

The data points that have positive marginal probability under λ^{k+1} but not under any of the previous priors are those of order $k+1$ for $\alpha_{j_1},\ldots,\alpha_{j_{k+1}}$. For any such point, the marginal probability is

$$\lambda^{k+1}(y(s)) \propto [\prod_{\ell=1}^{k+1} \Gamma(w_y(j_\ell;s))]/\Gamma(n(s)).$$

In this case, following the same steps as above, the result follows easily.

Continuing in this way until all possible data points are covered, we see that $U^*(y_{i_1},\ldots,y_{i_{n(s)}})$ is unique stepwise Bayes against that set of priors which implies that it is finitely admissible and, hence, is admissible and the proof is complete.

Remark 3.4:

From the above proof we see that, for any constant, say b, $bU^*(y_{i_1},\ldots,y_{i_{n(s)}})$ is an admissible estimator for $bU_p(y_1,\ldots,y_N)$.

Remark 3.5:

Note that $U^*(y_{i_1},\ldots,y_{i_{n(s)}})$ is a shrinkage estimator, i.e., an estimator of the form $aU_s(y_{i_1},\ldots,y_{i_{n(s)}})$ where $a \leq 1$. In fact, the shrinkage factor a is a function of k, say $\eta(k)$. For $k = 2$, it is easy to see that $\eta(2) \leq 1$. We conjecture that $\eta(k)$

is decreasing in k. But we were unable to prove that. However, computations of $\eta(k)$ for $N = 5, 10, 20, 50, 100$ and all values of $n(s) \leq N-1$ and $k \leq \min(n(s), N-n(s))$ support this conjecture. For some of those computations, please see Table 1.

In the rest of this section, we give some special cases.

Case (1): When $k = 1$ and $\xi(y_i) = y_i$ we get

$$U_p(y_1,\ldots,y_N) = [N]^{-1}\sum_{i=1}^{N} y_i \quad \text{and} \quad U^*(y_{i_1},\ldots,y_{i_{n(s)}}) = [n(s)]^{-1}\sum_{i\in s} y_i.$$

Hence, by Theorem 3.4, the sample mean is admissible for estimating the population mean. Moreover, by Remark 3.4, $[N/n(s)]\sum_{i\in s} y_i$ is an admissible estimator for $\sum_{i=1}^{N} y_i$. This result was first proved in Joshi (1965).

Case (2): When $k = 2$ and $\xi(y_i,y_j) = y_i-y_j$ we get

$$U_p(y_1,\ldots,y_N) = \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\ i<j}}^{N}(y_i-y_j)^2 = \frac{2}{(N-1)}\sum_{i=1}^{N}(y_i-\bar{y})^2$$

and

$$U^*(y_{i_1},\ldots,y_{i_{n(s)}}) = \frac{2(N+1)}{(N-1)(n(s))(n(s)+1)}\sum_{i\in s}\sum_{\substack{j\in s\\ i<j}}(y_i-y_j)^2$$

$$= \frac{2(N+1)((n(s)-1))}{(N-1)(n(s)+1)(n(s)-1)}\sum_{i\in s}(y_i-\bar{y}_s)^2$$

Table 1. Values of the shrinkage factor for various values of N, $n(s)$ and k where $n(s) \leq N-1$ and $k \leq \min(n(s), N-n(s))$

For N = 5

n(s) \ k	1	2
2	1	0.50
3	1	0.75
4	1	

For N = 10

n(s) \ k	1	2	3	4	5
2	1	0.407			
5	1	0.815	0.524	0.243	0.063
8	1	0.951			

For N = 20

n(s) \ k	1	2	3	4	5	6	7	8
4	1	0.663	0.270	0.052				
8	1	0.860	0.630	0.388	0.194	0.075	0.020	0.003
12	1	0.935	0.816	0.663	0.497	0.341	0.211	0.115
16	1	0.975	0.927	0.858				

For N = 50

n(s) \ k	2	4	6	8	10	12	14	16	18	20
10	0.852	0.373	0.077	0.006	0.00007					
20	0.942	0.696	0.399	0.175	0.056	0.013	0.002	0.0002	0.000006	0.00000004
30	0.974	0.852	0.667	0.467	0.289	0.157	0.074	0.030	0.010	0.003
40	0.990	0.941	0.860	0.753	0.631					

For N = 100

n(s) \ k	5	10	15	20	25	30	35	40
20	0.446	0.023	0.00007	0.0000000007				
60	0.875	0.546	0.240	0.072	0.014	0.002	0.0001	0.000005
80	0.951	0.797	0.587	0.378				

where $\bar{y}$ is the population mean and $\bar{y}_s$ is the sample mean. Theorem 3.4 and Remark 3.4 imply that $[\frac{(N+1)(n(s)-1)}{(N-1)(n(s)+1)}] \frac{1}{(n(s)-1)} \sum_{i\in s} (y_i-\bar{y}_s)^2$ is admissible for estimating the population variance, $[1/(N-1)] \sum_{i=1}^{N} (y_i-\bar{y})^2$. This estimator was first constructed in Ghosh and Meeden (1982).

3.2. In Nonparametric Problems

Let F be an unknown distribution that belongs to some nonparametric family of distributions, say Θ. For estimating with squared error loss some function of F, say $\tau(F)$, Meeden, Ghosh, and Vardeman (1984) have shown that admissible estimators for $\tau(F)$ can be obtained by considering only the subfamily of Θ consisting of all distribution functions which concentrate all their mass on a finite set of real numbers. In what follows, we will show that admissible pairs for $\tau(F)$ can be obtained in an analogue way.

Consider the decision problem specified by the class Θ of all distribution functions F for which $\tau(F) = \int\psi(t)dF(t)$ exist where $\psi(\cdot)$ is a specified continuous bounded function on the real line, a decision space D with generic element d, a squared error loss function and a collection of N random samples $X = \{\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_N\}$ from F where $\underset{\sim}{X}_\ell = (X_1,\ldots,X_\ell)$ and $\ell = 1,\ldots,N$. Let δ_ℓ, a continuous bounded function from $\mathcal{X}_\ell$, the sample space of $\underset{\sim}{X}_\ell$ to D, denote a typical decision function. Let $r_\ell(\delta_\ell;F)$ denote the risk function of δ_ℓ. In addition, we only consider those δ_ℓ's for which $r_\ell(\delta_\ell;F)$ is finite for all

$F \in \Theta$. We denote this class by Δ. Let $\Gamma^* = \{\underset{\sim}{\gamma}\}$ be a class of discrete probability measures defined on $\{1,\ldots,N\}$, i.e., $\underset{\sim}{\gamma} = (\gamma_1,\ldots,\gamma_N)$ where γ_ℓ is the probability of observing $\underset{\sim}{X}_\ell$ under $\underset{\sim}{\gamma}$.

For estimating $\tau(F)$, the interest is to obtain an admissible pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ relative to Γ^* where $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_N)$ and $\underset{\sim}{\gamma} \in \Gamma^*$. As we will soon show, admissible pairs for this problem can be obtained by obtaining admissible pairs for the following simpler problem:

Consider the problem of estimating $\tau(F)$ where F belongs to $\Theta(\alpha_1,\ldots,\alpha_r)$, the set of all distribution functions which concentrate all their mass on r distinct real numbers $\alpha_1,\ldots,\alpha_r$. In this case, $\underset{\sim}{X}_\ell$ is a random sample from a multinomial $(\nu_1,\ldots,\nu_r)$ population where $\nu_i = p(X_j = \alpha_i)$ for $i = 1,\ldots,r$ and $j = 1,\ldots,\ell$. For $\underset{\sim}{x}_\ell = (x_1,\ldots,x_\ell)$, a possible realization of $\underset{\sim}{X}_\ell$, let $w_i(\underset{\sim}{x}_\ell)$ be the number of x_j's equal to α_i for $i = 1,\ldots,r$. Note that $\Theta(\alpha_1,\ldots,\alpha_r)$ is equivalent to the r dimensional simplex

$$T = \{\underset{\sim}{\nu} = (\nu_1,\ldots,\nu_r)\colon \nu_i \geq 0 \text{ for } i = 1,\ldots,r \text{ and } \sum_{i=1}^{r} \nu_i = 1\}.$$

For $\underset{\sim}{\nu} \in T$ there exists a unique F corresponding to $\underset{\sim}{\nu}$ which we shall denote by $F_{\underset{\sim}{\nu}}$. Hence,

$$\tau(\underset{\sim}{\nu}) = \tau(F_{\underset{\sim}{\nu}}) = \sum_{i=1}^{r} \psi(\alpha_i)\nu_i.$$

By taking T to be the parameter space, admissible pairs for $\tau(\underset{\sim}{\nu})$ can be obtained by using the procedure given in Section 2.1.1. Now, the following theorem shows how admissibility in this simpler problem implies admissibility in the original problem.

Theorem 3.5:

Under the previous assumptions, if $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is admissible within the class Δ relative to Γ^* when $F \in \Theta(\alpha_1,\ldots,\alpha_r)$ for every choice of $\alpha_1,\ldots,\alpha_r$ for $r = 1,2,\ldots$ then, it is also admissible within the class Δ relative to Γ^* when $F \in \Theta$.

Proof:

Suppose $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is not admissible relative to Γ^* for the non-parametric problem then there exists a pair $(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^o)$ with $\underset{\sim}{\gamma}^o \in \Gamma^*$ and $\underset{\sim}{\delta}^o \in \Delta$ such that

$$r(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^o;F) \le r(\underset{\sim}{\gamma},\underset{\sim}{\delta};F) \quad \text{for all} \quad F \in \Theta$$

with strict inequality for at least one F, say F^*.

If F^* is a distribution which puts its mass on only finitely many points, say $(\alpha_1,\ldots,\alpha_r)$, then this will imply that $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ is not admissible relative to Γ^* for the simpler problem which is a contradiction.

So suppose F^* doesn't put all its mass on finitely many points, then there exists a sequence of distribution functions $\{F_n\}$ such that F_n converges completely to F^* and each F_n puts mass on only finitely many points. Now,

$$r(\underset{\sim}{\gamma}^o,\underset{\sim}{\delta}^o;F_n) = \sum_{\ell=1}^{N} \gamma_\ell^o \int [\delta_\ell^o - \tau(F_n)]^2 dF_n(t)$$

$$= \sum_{\ell=1}^{N} \gamma_\ell^o [\int \delta_\ell^{o^2} dF_n(t) - 2\tau(F_n)\int \delta_\ell^o dF_n(t) + \tau^2(F_n)\int dF_n(t)]$$

and by the Helly Bray theorem we have:

$$r(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};F_n) \rightarrow \sum_{\ell=1}^{N} \gamma_\ell^{o}[\int \delta_\ell^{o\,2} dF^*(t) - 2\tau(F^*) \int \delta_\ell^{o} dF^*(t) + \tau^2(F^*)\int dF^*(t)]$$

$$= r(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};F^*) .$$

Similarly, $r(\underset{\sim}{\gamma},\underset{\sim}{\delta};F_n) \rightarrow r(\underset{\sim}{\gamma},\underset{\sim}{\delta};F^*)$. Therefore, if $r(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};F^*) <$ $r(\underset{\sim}{\gamma},\underset{\sim}{\delta};F^*)$ then $r(\underset{\sim}{\gamma}^{o},\underset{\sim}{\delta}^{o};F_n) < r(\underset{\sim}{\gamma},\underset{\sim}{\delta};F_n)$ for some F_n which is a contradiction.

We now give an example to clarify the above idea.

Example:

For estimating with squared error loss $\tau(F) = \int\psi(t)dF(t)$ where $\psi(\cdot)$ is a continuous bounded function on the real line, we want to obtain an admissible pair $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ within the class Δ relative to Γ^* where $\Gamma^* = \{\underset{\sim}{\gamma}: \sum_{\ell=1}^{N} \ell\gamma_\ell \leq n\}$. Meeden, Ghosh, and Vardeman (1984) have obtained an admissible estimator for $\tau(F)$. We now use the same sequence of priors they used, in order to compute the Bayes risks and, hence, obtain an admissible pair for this problem. For real numbers $\alpha_1,\ldots,\alpha_r$, consider the multinomial problem with parameter space T. Let T_j be the subset of T consisting of all those $\underset{\sim}{\nu}$'s for which exactly j of the coordinates $\nu_1,\ldots,\nu_r$ are nonzeros. Consider the sequence of priors $g_1,\ldots,g_r$ such that g_1 puts mass $\frac{1}{r}$ on each of the r unit vectors belonging to T_1 and for $j > 1$ g_j is given by

$$g_j(\underset{\sim}{\nu}) \propto (\prod_{i:\nu_i>0} \nu_i)^{-1} \quad \text{for} \quad \underset{\sim}{\nu} \in T_j \quad j = 2,\ldots,r.$$

As in Meeden, Ghosh, and Vardeman (1984), given a random sample of size ℓ, $\frac{w_i(\underset{\sim}{x}_\ell)}{\ell}$ is a unique stepwise Bayes estimator against this sequence of priors and, hence, an admissible estimator of ν_i. Therefore, an admissible estimator for $\tau(\underset{\sim}{\nu})$ based on a random sample of size ℓ is

$$\delta_\ell = E(\tau|\underset{\sim}{x}_\ell) = \sum_{i=1}^{r} \psi(\alpha_i) \frac{w_i(\underset{\sim}{x}_\ell)}{\ell} = \sum_{j=1}^{\ell} \frac{\psi(x_j)}{\ell} = \bar{\psi}(\underset{\sim}{x}_\ell) .$$

We now compute the Bayes risk for a random sample of size ℓ.

$$R_\ell(\delta_\ell;g_1) = \sum_{\underset{\sim}{\nu}} \sum_{\underset{\sim}{x}_\ell} [\delta_\ell(\underset{\sim}{x}_\ell) - \tau(\underset{\sim}{\nu})]^2 f_{\underset{\sim}{\nu}}(\underset{\sim}{x}_\ell)g_1(\underset{\sim}{\nu})$$

$$= \sum_{\underset{\sim}{\nu}} \sum_{\underset{\sim}{x}_\ell} [\sum_{i=1}^{r} \psi(\alpha_i) \frac{w_i(\underset{\sim}{x}_\ell)}{\ell} - \sum_{i=1}^{r} \psi(\alpha_i)\nu_i]^2 f_{\underset{\sim}{\nu}}(\underset{\sim}{x}_\ell)g_1(\underset{\sim}{\nu})$$

$$= \sum_{j=1}^{r} [\psi(\alpha_j) - \psi(\alpha_j)]^2(1) \frac{1}{r} = 0 \quad \text{for all} \quad \ell .$$

Now, the Bayes risk under g_2 is

$$R_\ell(\delta_\ell;g_2) = \sum_{i=1}^{r} \underset{i<j}{\sum_{j=1}^{r}} \iint \sum_{\underset{\sim}{x}_\ell} (\delta_\ell - \tau)^2 f_{\underset{\sim}{\nu}}(\underset{\sim}{x}_\ell)g_2(\underset{\sim}{\nu})d\nu_i d\nu_j$$

$$= \sum_{i=1}^{r} \underset{i<j}{\sum_{j=1}^{r}} \iint \sum_{\underset{\sim}{x}_\ell} [\sum_{k=1}^{r} \psi(\alpha_k) \frac{w_k(\underset{\sim}{x}_\ell)}{\ell} - \sum_{k=1}^{r} \psi(\alpha_k)\nu_k]^2$$

$$\frac{\ell!}{w_i(\underset{\sim}{x}_\ell)!\,w_j(\underset{\sim}{x}_\ell)!}\,\nu_i^{w_i(\underset{\sim}{x}_\ell)}\,\nu_j^{w_j(\underset{\sim}{x}_\ell)}\,\frac{c}{\nu_i\nu_j}\,d\nu_i d\nu_j$$

where c is the normalizing constant of $g_2(\underset{\sim}{\nu})$. Note that from now on we will write w_i and w_j for $w_i(\underset{\sim}{x}_\ell)$ and $w_j(\underset{\sim}{x}_\ell)$. Hence,

$$R_\ell(\delta_\ell;g_2) = \sum_{\substack{i=1\\i<j}}^{r}\sum_{j=1}^{r}\iint\sum_{\underset{\sim}{x}_\ell}[\psi(\alpha_i)\frac{w_i}{\ell}+\psi(\alpha_j)\frac{w_j}{\ell}-\psi(\alpha_i)\nu_i-\psi(\alpha_j)\nu_j]^2$$

$$\frac{\ell!}{w_i!w_j!}\,\nu_i^{w_i}\nu_j^{w_j}\,\frac{c}{\nu_i\nu_j}\,d\nu_i d\nu_j$$

$$= \sum_{\substack{i=1\\i<j}}^{r}\sum_{j=1}^{r}\iint\sum_{\underset{\sim}{x}_\ell}\{\frac{1}{\ell}\,\psi(\alpha_i)[w_i-\ell\nu_i]+\frac{1}{\ell}\,\psi(\alpha_j)[w_j-\ell\nu_j]\}^2$$

$$\frac{\ell!}{w_i!w_j!}\,\nu_i^{w_i}\nu_j^{w_j}\,\frac{c}{\nu_i\nu_j}\,d\nu_i d\nu_j$$

$$= \sum_{\substack{i=1\\i<j}}^{r}\sum_{j=1}^{r}\iint\{\frac{\psi^2(\alpha_i)}{\ell^2}\sum_{\underset{\sim}{x}_\ell}[w_i-\ell\nu_i]^2\,\frac{\ell!}{w_i!w_j!}\,\nu_i^{w_i}\nu_j^{w_j}$$

$$+\frac{\psi^2(\alpha_j)}{\ell^2}\sum_{\underset{\sim}{x}_\ell}[w_j-\ell\nu_j]^2\,\frac{\ell!}{w_i!w_j!}\,\nu_i^{w_i}\nu_j^{w_j}$$

$$+\frac{2\psi(\alpha_i)\psi(\alpha_j)}{\ell^2}\sum_{\underset{\sim}{x}_\ell}[w_i-\ell\nu_i][w_j-\ell\nu_j]\,\frac{\ell!}{w_i!w_j!}\,\nu_i^{w_i}\nu_j^{w_j}\}\,\frac{c}{\nu_i\nu_j}\,d\nu_i d\nu_j$$

$$= \sum_{\substack{i=1 \\ i<j}}^{r} \sum_{j=1}^{r} \iint \{ \frac{\psi^2(\alpha_i)}{\ell^2} [\ell\nu_i\nu_j] + \frac{\psi^2(\alpha_j)}{\ell^2} [\ell\nu_i\nu_j]$$

$$+ \frac{2\psi(\alpha_i)\psi(\alpha_j)}{\ell^2} [-\ell\nu_i\nu_j]\} \frac{c}{\nu_i\nu_j} d\nu_i d\nu_j$$

$$= \sum_{\substack{i=1 \\ i<j}}^{r} \sum_{j=1}^{r} \iint \frac{1}{\ell} [\psi(\alpha_i) - \psi(\alpha_j)]^2 \; c \; d\nu_i d\nu_j$$

$$= (c/\ell) \sum_{\substack{i=1 \\ i<j}}^{r} \sum_{j=1}^{r} [\psi(\alpha_i) - \psi(\alpha_j)]^2 .$$

Now, for any $\underset{\sim}{\gamma} \in \Gamma^*$, the Bayes risk of $(\underset{\sim}{\gamma}, \underset{\sim}{\delta})$ under g_2 is

$$R(\underset{\sim}{\gamma}, \underset{\sim}{\delta}; g_2) = \sum_{\ell=1}^{N} \gamma_\ell R_\ell(\delta_\ell; g_2)$$

$$= \sum_{\ell=1}^{N} \gamma_\ell (c/\ell) \sum_{1\le i<j\le r} \sum [\psi(\alpha_i) - \psi(\alpha_j)]^2$$

$$= c \sum_{1\le i<j\le r} \sum [\psi(\alpha_i) - \psi(\alpha_j)]^2 \sum_{\ell=1}^{N} (\gamma_\ell/\ell)$$

$$\ge c \sum_{1\le i<j\le r} \sum [\psi(\alpha_i) - \psi(\alpha_j)]^2 \; [1/ \sum_{\ell=1}^{N} \ell\gamma_\ell]$$

$$\ge c \sum_{1\le i<j\le r} \sum [\psi(\alpha_i) - \psi(\alpha_j)]^2 \; [1/n]$$

$$= R_n(\delta_n; g_2)$$

where the sign of equality holds if and only if $\ell = n$ for all possible random samples. Therefore,

$$\inf_{\underset{\sim}{\gamma}\in\Gamma^*} R(\underset{\sim}{\gamma},\underset{\sim}{\delta};g_2) = R(\underset{\sim}{\gamma}',\underset{\sim}{\delta};g_2)$$

where $\underset{\sim}{\gamma}'$ is the probability measure that chooses with probability one, the random sample of size n. Hence, $(\underset{\sim}{\gamma}',\underset{\sim}{\delta})$ is admissible within Δ relative to Γ^*.

3.3. A Uniform Admissibility Duality Between Nonparametric and Finite Population Sampling

In this section, we show that there is a Bayes risk duality between nonparametric and finite population sampling problems which implies that under some conditions, uniform admissibility results in finite population sampling can be obtained by considering only the nonparametric problem and vice versa.

We now briefly represent the two decision problems given in Sections 3.1 and 3.2.

The Nonparametric Problem:

Let Θ denote the class of all distribution functions F for which $\tau(F) = \int\psi(t)dF(t)$ exist where $\psi(\cdot)$ is a specified continuous bounded function on the real line. Let $X = \{\underset{\sim}{X}\colon \underset{\sim}{X} = (X_1,\ldots,X_{n(x)}),$ $n(x) = 1,\ldots,N\}$ be a collection of N random samples from F. Let $\Gamma^* = \{\underset{\sim}{\gamma}\colon \underset{\sim}{\gamma} = (\gamma_1,\ldots,\gamma_N)\}$ be a class of discrete probability measures defined on $\{1,\ldots,N\}$. For estimating $\tau(F)$ with squared error loss, the interest is to know which $\underset{\sim}{X}$ should be observed and which estimator should be used, i.e., to characterize the admissible pairs $(\underset{\sim}{\gamma},\underset{\sim}{\delta})$ relative to Γ^* where $\underset{\sim}{\delta} = (\delta_1,\ldots,\delta_N)$ and $\delta_i \in \Delta$ where Δ is as defined in Section 3.2.

The Sampling Problem:

In a population of size N, let $y = (y_1,\ldots,y_N) \in \mathbb{R}^N$ be the parameter of interest and S be the set of all possible samples from this population. For $s \in S$ let $n(s)$ be the size of s where $n(s) = 1,2,\ldots,N$ and $y(s) = (y_{i_1},\ldots,y_{i_{n(s)}})$ be the values in this sample. Define Γ^*, a class of discrete probability measures on S, in the same way as in the nonparametric problem. Let F_y be the distribution function which assigns mass $\frac{1}{N}$ to each component y_i of y. For estimating with squared error loss, $\tau(y)$ where

$$\tau(y) = \tau(F_y) = \int\psi(t)dF_y(t) = \sum_{i=1}^{N} \psi(y_i)/N ,$$

the interest is to characterize the uniformly admissible pairs $(\underset{\sim}{\gamma},\underset{\sim}{\delta}^*)$ relative to Γ^* where $\underset{\sim}{\delta}^* = (\delta_1^*,\ldots,\delta_M^*)$ and δ_ℓ^* is the decision function to be used in connection with the ℓ^{th} sample, $\ell = 1,\ldots,M$ and M is the number of elements in S. Note that, while there are many samples of size ℓ, $\ell = 1,\ldots,N$ for the sampling problem, there is only one random sample of size ℓ for the nonparametric problem.

For a typical random sample $\underset{\sim}{X}$ of size $n(x)$ and a typical sample s of size $n(s)$, Meeden, Ghosh, and Vardeman (1984) have shown, by reducing those problems to simpler ones, that there is a duality between admissible estimators in the two problems. Using this duality we will show that there is a Bayes risk duality between the two problems.

Now, for the purpose of completeness, we first represent the proof, given by Meeden, Ghosh, and Vardeman (1984), of the duality between admissible estimators in the two problems.

The Nonparametric Problem:

As noted by Meeden, Ghosh, and Vardeman (1984), to prove that an estimator is admissible it is enough to show that it is admissible for the multinomial problem with parameter space $\Theta(\alpha_1,\ldots,\alpha_r)$ for every choice of $\alpha_1,\ldots,\alpha_r$. Let G be a prior distribution over $T = \{\underset{\sim}{\nu}\colon\ \underset{\sim}{\nu} = (\nu_1,\ldots,\nu_r);\ \nu_i \geq 0,\ \sum_{i=1}^{r} \nu_i = 1\}$. If $\underset{\sim}{x} = (x_1,\ldots,x_{n(x)})$ is a possible set of outcomes for the random sample $\underset{\sim}{X}$ then let $w_i(\underset{\sim}{x})$ be the number of x_j's equal to α_i for $i = 1,\ldots,r$. Hence, the Bayes estimate of ν_j against G is

$$E_G(\nu_j|\underset{\sim}{x}) = \frac{\int\cdots\int \nu_j \prod_{i=1}^{r} \nu_i^{w_i(\underset{\sim}{x})}\, dG(\nu_1,\ldots,\nu_r)}{\int\cdots\int \prod_{i=1}^{r} \nu_i^{w_i(\underset{\sim}{x})}\, dG(\nu_1,\ldots,\nu_r)} = p(\alpha_j|\underset{\sim}{x},\ G) \tag{3.26}$$

where $p(\alpha_j|\underset{\sim}{x}\ ,\ G)$ is the G posterior probability that an additional observation takes the value α_j. From this, it follows that the Bayes estimate of τ against G is

$$E_G(\tau|\underset{\sim}{x}) = \sum_{i=1}^{r} \psi(\alpha_i)\, p(\alpha_i|\underset{\sim}{x}, G) \tag{3.27}$$

The Sampling Problem:

As we have seen from Section 3.1, admissible estimators for $\tau(y)$ can be obtained, using the idea of finite admissibility, from the simpler problem with parameter space $\bar{\Theta}(\alpha_1,\ldots,\alpha_r)$. For $y \in \bar{\Theta}(\alpha_1,\ldots,\alpha_r)$ let $w_j(y)$ be the number of y_i's equal to α_j, and $w_j(y(s))$ be the number of y_i's with $i \in s$ equal to α_j. Let G be as above and define the prior distribution G^* over $\bar{\Theta}(\alpha_1,\ldots,\alpha_r)$ as follows:

$$G^*(y) \propto \int\cdots\int \prod_{i=1}^{r} \nu_i^{w_i(y)}\, dG(\nu_1,\ldots,\nu_r) \quad \text{for} \quad y \in \bar{\Theta}(\alpha_1,\ldots,\alpha_r).$$

Hence, the Bayes estimate of τ against G^* is

$$E_{G^*}(\tau|y(s)) = \frac{1}{N}\left\{\sum_{j\in s}\psi(y_j) + \sum_{j\notin s} E_{G^*}(\psi(y_j)|y(s))\right\}$$

$$= \frac{1}{N}\left\{\sum_{i=1}^{r}\psi(\alpha_i)w_i(y(s)) + (N-n(s))E_{G^*}(\psi(y_{j'})|y(s))\right\}$$

where $j' \notin s$. Now, it is easy to see that the posterior distribution of G^* assigns probability $p(\alpha_i|y(s), G)$ to the event that an unobserved $y_{j'}$ takes on the value α_i. Hence,

$$E_{G*}(\tau|y(s)) = \frac{1}{N}\{\sum_{i=1}^{r}\psi(\alpha_i)w_i(y(s)) + (N-n(s))\sum_{i=1}^{r}\psi(\alpha_i)p(\alpha_i|y(s),G)\}$$

$$= \frac{1}{N}\{\sum_{i=1}^{r}\psi(\alpha_i)w_i(y(s)) + (N-n(s))\,E_G(\tau|y(s))\}.$$

Letting $\delta^*(y(s)) = E_{G*}(\tau|y(s))$ and $\delta(\underset{\sim}{x}) = E_G(\tau|\underset{\sim}{x})$ we get

$$\delta^*(y(s)) = \frac{1}{N}\sum_{i\in s}\psi(y_i) + \frac{(N-n(s))}{N}\,\delta(y(s)). \tag{3.28}$$

Equation (3.28) says that if δ is unique Bayes (unique stepwise Bayes) against G (a sequence of priors) for the multinomial problem with parameter space T then δ is admissible. Moreover, δ^* is unique Bayes (unique stepwise Bayes) against G^* (a sequence of priors) for the finite population sampling problem with parameter space $\bar{\Theta}(\alpha_1,\ldots,\alpha_r)$ and, hence, is admissible under any design. As noted by Meeden, Ghosh, and Vardeman (1984), if this result holds for every choice of $(\alpha_1,\ldots,\alpha_r)$ then δ and δ^* are admissible for the original problems as well.

For more details about this duality, please see Meeden, Ghosh, and Vardeman (1984).

We now use this duality to give a corresponding Bayes risk duality.

A Bayes Risk Duality:

Recall that for the multinomial problem, we have

$$\tau(\underset{\sim}{\nu}) = \sum_{i=1}^{r}\psi(\alpha_i)\nu_i \quad \text{and}$$

$$\delta(\underset{\sim}{x}) = \sum_{i=1}^{r} \psi(\alpha_i) E_G(\nu_i|\underset{\sim}{x}).$$

Therefore, using the prior density $g(\nu_1,\ldots,\nu_r)$, the posterior distribution, say $\Phi(\underset{\sim}{\nu}|\underset{\sim}{x})$, is

$$\Phi(\underset{\sim}{\nu}|\underset{\sim}{x}) \propto \prod_{i=1}^{r} \nu_i^{w_i(x)} \quad g(\nu_1,\ldots,\nu_r).$$

Hence, the posterior risk, say $\rho(\underset{\sim}{x})$, is

$$\rho(\underset{\sim}{x}) = \int\ldots\int [\tau(\underset{\sim}{\nu})-\delta(\underset{\sim}{x})]^2 \; \phi(\underset{\sim}{\nu}|\underset{\sim}{x}) \; d\nu_1,\ldots,d\nu_r$$

$$= \int\ldots\int \Big[\sum_{i=1}^{r} \psi(\alpha_i)\nu_i - \sum_{i=1}^{r} \psi(\alpha_i) E_G(\nu_i|\underset{\sim}{x})\Big]^2 \; \phi(\underset{\sim}{\nu}|\underset{\sim}{x}) \; d\nu_1,\ldots,d\nu_r$$

$$= \int\ldots\int \Big\{\sum_{i=1}^{r} \psi(\alpha_i)[\nu_i - E_G(\nu_i|\underset{\sim}{x})]\Big\}^2 \; \phi(\underset{\sim}{\nu}|\underset{\sim}{x}) \; d\nu_1,\ldots,d\nu_r$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i) \mathrm{Var}_G(\nu_i|\underset{\sim}{x}) + 2 \sum_{1\le i<j\le r}\sum \psi(\alpha_i)\psi(\alpha_j) \; \mathrm{Cov}_G(\nu_i,\nu_j|\underset{\sim}{x}).$$

Let $\phi_1(\underset{\sim}{x})$ denote the marginal of $\underset{\sim}{X}$, then the Bayes risk is

$$R_{\underset{\sim}{X}}(\delta(\underset{\sim}{x}),G) = \sum_{i=1}^{r} \psi^2(\alpha_i) \sum_{\underset{\sim}{x}} \mathrm{Var}_G(\nu_i|\underset{\sim}{x})\phi_1(\underset{\sim}{x})$$

$$+ 2 \sum_{1\le i<j\le r}\sum \psi(\alpha_i)\psi(\alpha_j) \sum_{\underset{\sim}{x}} \mathrm{Cov}_G(\nu_i,\nu_j|\underset{\sim}{x})\phi_1(\underset{\sim}{x}). \qquad (3.29)$$

Also, recall that for the simpler sampling problem we have

$$\tau(y) = \frac{1}{N}\sum_{i=1}^{N}\psi(y_i) \quad \text{and}$$

$$\delta^*(y(s)) = \frac{1}{N}\sum_{i\in s}\psi(y_i) + \frac{[N-n(s)]}{N}\,\delta(y(s)).$$

Let $g_1^*(y|y(s))$ and $g_2^*(y(s))$ be the posterior and the marginal distributions, respectively. Then, the Bayes risk is

$$R_s(\delta^*,G^*) = \sum_{y(s)}\sum_{y}[\delta^*-\tau(y)]^2\, g_1^*(y|y(s))g_2^*(y(s))$$

$$= \sum_{y(s)}\sum_{y}\Big[\frac{1}{N}\sum_{i\in s}\psi(y_i) + \frac{(N-n(s))}{N}\,\delta(y(s)) - \frac{1}{N}\sum_{i=1}^{N}\psi(y_i)\Big]^2$$

$$g_1^*(y|y(s))g_2^*(y(s))$$

$$= \sum_{y(s)}\sum_{y}\Big[\frac{1}{N}\sum_{i\notin s}(\psi(y_i)-\delta(y(s)))\Big]^2\, g_1^*(y|y(s))g_2^*(y(s))$$

$$= \sum_{y(s)}\sum_{y}\Big\{\frac{1}{N}\sum_{i\notin s}[\psi(y_i)-E_{G^*}(\psi(y_i)|y(s))]\Big\}^2\, g_1^*(y|y(s))g_2^*(y(s))$$

$$= \frac{1}{N^2}\sum_{y(s)}\sum_{y}\Big\{\sum_{i\notin s}[\psi(y_i)-E_{G^*}(\psi(y_i)|y(s))]^2$$

$$+ \sum_{\substack{i\notin s\\ i\neq j}}\sum_{j\notin s}[\psi(y_i)-E_{G^*}(\psi(y_i)|y(s))][\psi(y_j)-E_{G^*}(\psi(y_j)|y(s))]\Big\}$$

$$g_1^*(y|y(s))g_2^*(y(s))$$

$$= \frac{1}{N^2} \sum_{y(s)} [(N-n(s))\, \mathrm{Var}_{G*}(\psi(y_i)|y(s))$$

$$+ (N-n(s))(N-n(s)-1)\mathrm{Cov}_{G*}(\psi(y_i),\psi(y_j)|y(s))]g_2^*(y(s)) \tag{3.30}$$

Note that

$$\mathrm{Var}_{G*}(\psi(y_i)|y(s)) = E_{G*}(\psi^2(y_i)|y(s)) - E^2_{G*}(\psi(y_i)|y(s))$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i)E_G(\nu_i|y(s)) - [\sum_{i=1}^{r} \psi(\alpha_i)E_G(\nu_i|y(s)]^2$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i)E_G(\nu_i|y(s)) - \sum_{i=1}^{r} \psi^2(\alpha_i)E^2_G(\nu_i|y(s))$$

$$- 2 \sum\sum_{1\le i<j\le r} \psi(\alpha_i)\psi(\alpha_j)E_G(\nu_i|y(s))E_G(\nu_j|y(s))$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i)E_G(\nu_i|y(s))[1-E_G(\nu_i|y(s))]$$

$$- 2 \sum\sum_{1\le i<j\le r} \psi(\alpha_i)\psi(\alpha_j)E_G(\nu_i|y(s))E_G(\nu_j|y(s))$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i)E_G(\nu_i|y(s))E_G((1-\nu_i)|y(s))$$

$$- 2 \sum\sum_{1\le i<j\le r} \psi(\alpha_i)\psi(\alpha_j)E_G(\nu_i|y(s))E_G(\nu_j|y(s))$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i) E_G(\nu_i | y(s)) E_G(\sum_{\substack{j=1 \\ j \neq i}}^{r} \nu_j | y(s))$$

$$- 2 \sum_{1 \le i < j \le r} \sum \psi(\alpha_i)\psi(\alpha_j) E_G(\nu_i | y(s)) E_G(\nu_j | y(s))$$

$$= \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \neq j}}^{r} \psi^2(\alpha_i) E_G(\nu_i | y(s)) E_G(\nu_j | y(s))$$

$$- 2 \sum_{1 \le i < j \le r} \sum \psi(\alpha_i)\psi(\alpha_j) E_G(\nu_i | y(s)) E_G(\nu_j | y(s))$$

$$= \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \neq j}}^{r} E_G(\nu_i | y(s)) E_G(\nu_j | y(s)) [\psi^2(\alpha_i) - \psi(\alpha_i)\psi(\alpha_j)]$$

$$= \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \neq j}}^{r} [-\mathrm{Cov}_G(\nu_i, \nu_j | y(s)) + E_G(\nu_i \nu_j | y(s))]$$

$$[\psi^2(\alpha_i) - \psi(\alpha_i)\psi(\alpha_j)]$$

$$= - \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \neq j}}^{r} \psi^2(\alpha_i) \mathrm{Cov}_G(\nu_i, \nu_j | y(s))$$

$$+ \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \neq j}}^{r} \psi(\alpha_i)\psi(\alpha_j) \mathrm{Cov}_G(\nu_i, \nu_j | y(s))$$

$$+ \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \neq j}}^{r} E_G(\nu_i \nu_j | y(s)) [\psi^2(\alpha_i) - \psi(\alpha_i)\psi(\alpha_j)]$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i)\mathrm{Cov}_G(\nu_i, -\sum_{\substack{j=1\\ j\neq i}}^{r} \nu_j | y(s))$$

$$+ 2 \sum_{1\leq i<j\leq r} \sum \psi(\alpha_i)\psi(\alpha_j)\mathrm{Cov}_G(\nu_i,\nu_j|y(s))$$

$$+ \sum_{\substack{i=1\\ i\neq j}}^{r} \sum_{j=1}^{r} E_G(\nu_i\nu_j|y(s)) [\psi^2(\alpha_i) - \psi(\alpha_i)\psi(\alpha_j)]$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i)\mathrm{Cov}_G(\nu_i, \nu_i-1|y(s))$$

$$+ 2 \sum_{1\leq i<j\leq r} \sum \psi(\alpha_i)\psi(\alpha_j)\mathrm{Cov}_G(\nu_i,\nu_j|y(s))$$

$$+ \sum_{\substack{i=1\\ i\neq j}}^{r} \sum_{j=1}^{r} E_G(\nu_i\nu_j|y(s)) [\psi^2(\alpha_i) - \psi(\alpha_i)\psi(\alpha_j)]$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i)\mathrm{Var}_G(\nu_i|y(s)) + 2 \sum_{1\leq i<j\leq r} \sum \psi(\alpha_i)\psi(\alpha_j)\mathrm{Cov}_G(\nu_i,\nu_j|y(s))$$

$$+ \sum_{\substack{i=1\\ i\neq j}}^{r} \sum_{j=1}^{r} E_G(\nu_i\nu_j|y(s)) [\psi^2(\alpha_i) - \psi(\alpha_i)\psi(\alpha_j)]. \quad (3.31)$$

Also,

$$\mathrm{Cov}_{G*}(\psi(y_i),\psi(y_j)|y(s)) = E_{G*}(\psi(y_i)\psi(y_j)|y(s))$$

$$- E_{G*}(\psi(y_i)|y(s))E_{G*}(\psi(y_j)|y(s))$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{r} \psi(\alpha_i)\psi(\alpha_j) E_G(\nu_i \nu_j | y(s))$$

$$- [\sum_{i=1}^{r} \psi(\alpha_i) E_G(\nu_i | y(s))][\sum_{j=1}^{r} \psi(\alpha_j) E_G(\nu_j | y(s))]$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i) E_G(\nu_i^2 | y(s))$$

$$+ 2 \sum_{1 \le i < j \le r} \sum \psi(\alpha_i)\psi(\alpha_j) E_G(\nu_i \nu_j | y(s))$$

$$- \sum_{i=1}^{r} \psi^2(\alpha_i) E_G^2(\nu_i | y(s))$$

$$- 2 \sum_{1 \le i < j \le r} \sum \psi(\alpha_i)\psi(\alpha_j) E_G(\nu_i | y(s)) E_G(\nu_j | y(s))$$

$$= \sum_{i=1}^{r} \psi^2(\alpha_i) Var_G(\nu_i | y(s))$$

$$+ 2 \sum_{1 \le i < j \le r} \sum \psi(\alpha_i)\psi(\alpha_j) Cov_G(\nu_i, \nu_j | y(s)). \quad (3.32)$$

Now, substitute with (3.31) and (3.32) in (3.30) we get

$$R_s(\hat{\delta}^*, G^*) = \frac{(N-n(s))}{N^2} \sum_{y(s)} \sum_{\substack{i=1 \\ i \ne j}}^{r} \sum_{j=1}^{r} E_G(\nu_i \nu_j | y(s)) [\psi^2(\alpha_i) - \psi(\alpha_i)\psi(\alpha_j)] g_2^*(y(s))$$

$$+ (\frac{N-n(s)}{N})^2 \sum_{y(s)} [\sum_{i=1}^{r} \psi^2(\alpha_i) Var_G(\nu_i | y(s))$$

$$+ 2 \sum_{1 \le i<j \le r} \sum \psi(\alpha_i)\psi(\alpha_j)\mathrm{Cov}_G(\nu_i,\nu_j|y(s))]g_2^*(y(s))$$

$$= \frac{(N-n(s))}{N^2} \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \ne j}}^{r} [\psi^2(\alpha_i)-\psi(\alpha_i)\psi(\alpha_j)]$$

$$\sum_{y(s)} E_G(\nu_i\nu_j|y(s))g_2^*(y(s))$$

$$+ (\frac{N-n(s)}{N})^2 [\sum_{i=1}^{r} \psi^2(\alpha_i) \sum_{y(s)} \mathrm{Var}_G(\nu_i|y(s))g_2^*(y(s))$$

$$+ 2 \sum_{1 \le i<j \le r} \sum \psi(\alpha_i)\psi(\alpha_j) \sum_{y(s)} \mathrm{Cov}_G(\nu_i,\nu_j|y(s))g_2^*(y(s))]$$

$$= \frac{(N-n(s))}{N^2} \sum_{i=1}^{r} \sum_{\substack{j=1 \\ i \ne j}}^{r} \{\psi^2(\alpha_i)-\psi(\alpha_i)\psi(\alpha_j)\} E_G(\nu_i\nu_j)$$

$$+ (\frac{N-n(s)}{N})^2 R_{y(s)}(\delta,G). \tag{3.33}$$

Now, considering the simpler problems, if δ is unique Bayes against G then δ^* is unique Bayes against G^*. Therefore, using Corollary 2.1, uniform admissibility can be studied in the two problems using the duality given in (3.33) provided that the assumption given in the Corollary (or the alternative assumption given in Remark 2.2) is satisfied for both X and S. If these uniform admissibility results hold under every choice of $(\alpha_1,\ldots,\alpha_r)$, then those results hold for the original problems as well.

In the special case, when the interest is to study uniform admissibility in both problems relative to the class of designs of fixed sample size n (in the nonparametric problem, this class consists of only one design namely, the design which picks the random sample of size n with probability one) we see that the above duality leads to the following result whose proof follows immediately from the above discussion.

Theorem 3.6:

For estimating τ with squared error loss, if δ is unique Bayes then $(\underset{\sim}{\gamma}, \underset{\sim}{\delta}^*)$ is uniformly admissible relative to the class of designs of fixed sample size n provided that the assumption in Corollary 2.1 (or the alternative assumption given in Remark 2.2) is satisfied for S_n.

4. REFERENCES

Alam, K. 1979. Estimation of multinomial probabilities. Ann. Statist. 7:282-283.

Basu, D. 1969. Role of the sufficiency and likelihood principles in sample survey theory. Sankhyā 31:441-454.

Basu, D. 1971. An essay on the logical foundations of survey sampling, part one. Foundations of Statistical Inference, edited by V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto.

Brown, L. D. 1981. A complete class theorem for statistical problems with finite sample spaces. Ann. Statist. 9:1289-1300.

Cassel, C. M., Sarndal, C. E., and Wretman, J. H. 1977. Foundations of Inference in Survey Sampling. John Wiley & Sons, New York.

Chaudhuri, A. 1978. On estimating the variance of a finite population. Metrika 25:65-76.

Cohen, M. and Kuo, J. 1984. The admissibility of the empirical distribution function. To appear in Ann. Statist.

DeGroot, M. H. 1970. Optimal Statistical Decisions. McGraw Hill, New York.

DeGroot, M. H. and Rao, M. M. 1963. Bayes estimation with convex loss. Ann. Math. Statist. 34:839-846.

Ericson, W. A. 1970. On a class of uniformly admissible estimators of a finite population total. Ann. Math. Statist. 41:1369-1372.

Ferguson, T. S. 1967. Mathematical Statistics, A Decision Theoretic Approach. Academic Press, New York.

Ghosh, M. and Meeden, G. 1982. Estimation of the variance in finite population sampling. To appear in Sankhyā, B.

Godambe, V. P. 1969. Admissibility and Bayes estimation in sampling finite populations V. Ann. Math. Statist. 40:672-676.

Hsuan, F. 1979. A stepwise Bayesian procedure. Ann. Statist. 4:860-868.

Joshi, V. M. 1965. Admissibility and Bayes estimation in sampling finite populations, II and III. Ann. Math. Statist. 36:1723-1742.

Joshi, V. M. 1966. Admissibility and Bayes estimation in sampling finite populations, IV. Ann. Math. Statist. 37:1658-1670.

Joshi, V. M. and Godambe, V. P. 1965. Admissibility and Bayes estimation in sampling finite populations, I. Ann. Math. Statist. 36:1707-1722.

Lehmann, E. L. 1983. Theory of Point Estimation. John Wiley & Sons, New York.

Meeden, G. and Ghosh, M. 1981. Admissibility in finite problems. Ann. Statist. 9:846-852.

Meeden, G. and Ghosh, M. 1982. On the admissibility and uniform admissibility of ratio type estimates. To appear in the Proceedings of the Golden Jubilee Conference of the Indian Statistical Institute.

Meeden, G. and Ghosh, M. 1983. Choosing between experiments: Applications to finite population sampling. Ann. Statist. 11:296-305.

Meeden, G., Ghosh, M., and Vardeman, S. 1984. Some admissible nonparametric and related finite population sampling estimators. (Unpublished manuscript.) Department of Statistics. I.S.U.

Randles, R. H. and Wolfe, D. A. 1979. Introduction to the Theory of Nonparametric Statistics. John Wiley & Sons, New York.

Scott, A. J. 1975. On admissibility and uniform admissibility in finite population sampling. Ann. Statist. 3:489-491.

Vardeman, S. and Meeden, G. 1983a. Admissible estimators in finite population sampling employing various types of prior information. J. Statist. Planning Inf. 7(4):329-341.

Vardeman, S. and Meeden, G. 1983b. Admissible estimators of the population total using trimming and Winsorization. Statistics and Prob. Letters 1:317-321.

Vardeman, S. and Meeden, G. 1984. Admissible estimators for the total of a stratified population that employ prior information. To appear in Ann. Statist.

Zacks, Shelley. 1969. Bayes sequential designs of fixed size samples from finite populations. J. Amer. Statist. Assoc. 64:1342-1349.

5. ACKNOWLEDGMENTS

I wish to express my sincere appreciation and gratitude to Professor Glen Meeden, to whom I am greatly indebted. Professor Meeden suggested the topic of this dissertation and he was a constant source of guidance and help during the course of this study. Without his help I would have been, certainly, unable to do this work. I am grateful to him, also, for his constructive comments which result in many improvements of the manuscript.

Special thanks and gratitude go to Professor Dean Isaacson for his guidance the first year of my study at Iowa State University and for his encouragement from that time on.

I am also grateful to all members of my graduate committee.

The typing of the manuscript is the superb job of Mrs. Sharon Shepard and I am thankful for that.

Thanks are also due to the Statistics Department, College of Economics and Political Science, Cairo University for giving me the chance to study at Iowa State University.

6. APPENDIX

Optimization Problems:

Let $\Psi_1(\cdot)$ and $\Psi_2(\cdot)$ be two functions defined on the set $N^* = \{1,2,\ldots,N\}$. Let $\mathcal{P}$ be the class of all possible probability measures defined on N^*. Let n be a value that belongs to the range of $\Psi_1(\cdot)$. Consider the following four optimization problems:

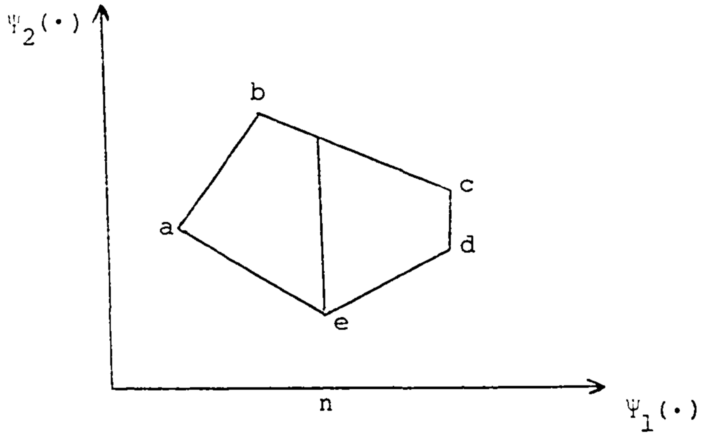
$$(1) \quad \min_{p\in\mathcal{P}} \sum_{i=1}^{N} \Psi_2(i)p(i) \quad \text{subject to} \quad \sum_{i=1}^{N} \Psi_1(i)p(i) \le n$$

$$(2) \quad \min_{p\in\mathcal{P}} \sum_{i=1}^{N} \Psi_2(i)p(i) \quad \text{subject to} \quad \sum_{i=1}^{N} \Psi_1(i)p(i) \ge n$$

$$(3) \quad \max_{p\in\mathcal{P}} \sum_{i=1}^{N} \Psi_2(i)p(i) \quad \text{subject to} \quad \sum_{i=1}^{N} \Psi_1(i)p(i) \le n$$

$$(4) \quad \max_{p\in\mathcal{P}} \sum_{i=i}^{N} \Psi_2(i)p(i) \quad \text{subject to} \quad \sum_{i=i}^{N} \Psi_1(i)p(i) \ge n$$

In general, the answer to any of the above problems can be obtained by graphing the convex set, say $\mathcal{R}$, generated by the points $(\Psi_1(i), \Psi_2(i))$ for all $i = 1,\ldots,N$. For example, suppose the graph of $\mathcal{R}$ is as follows:



Now, by looking at the lower boundary of R, we see that the first two problems are solved by taking p such that $p(n) = 1$ which gives the minimum value $\Psi_2(n)$ [this solution is represented by the point e in the graph]. On the other hand, the solution to the last two problems can be obtained by looking at the upper boundary of R. For instance, suppose that the points b and c correspond to $(\Psi_1(i_1), \Psi_2(i_1))$ and $(\Psi_1(i_2), \Psi_2(i_2))$ respectively, then we see that problem (3) is solved by taking p such that $p(i_1) = 1$ which gives the maximum value $\Psi_2(i_1)$, while problem (4) is solved by taking p such that $p(i_1) + p(i_2) = 1$ and $\Psi_1(i_1)p(i_1) + \Psi_1(i_2)p(i_2) = n$.

The graph of the set R, when N is large, can be done using a computer in the following manner: Define the sets R_1 and R_2 as follows:

$$R_1 = \{(\Psi_1(i_1), \Psi_2(i_1)): \ \Psi_1(i_1) = \min_i \Psi_1(i)\}$$

$$R_2 = \{(\Psi_1(i_N), \Psi_2(i_N)): \ \Psi_1(i_N) = \max_i \Psi_1(i)\}$$

Note that if $i_1(i_N)$ is unique then $\mathcal{R}_1(\mathcal{R}_2)$ consists of only one point. Now, to graph the upper boundary of $\mathcal{R}$ let:

$$\Psi_2(i_1^*) = \max_{i_1:(\Psi_1(i_1),\Psi_2(i_1))\in\mathcal{R}_1} \Psi_2(i_1) \quad \text{and} \quad \Psi_2(i_N^*) = \max_{i_N:(\Psi_1(i_N),\Psi_2(i_N))\in\mathcal{R}_2} \Psi_2(i_N)$$

Now, given i_1^* let i_2^* be such that

$$\frac{\Psi_2(i_2^*)-\Psi_2(i_1^*)}{\Psi_1(i_2^*)-\Psi_1(i_1^*)} = \max_{i:\Psi_1(i)>\Psi_1(i_1^*)} \frac{\Psi_2(i)-\Psi_2(i_1^*)}{\Psi_1(i)-\Psi_1(i_1^*)} \tag{6.1}$$

and among all i_2^*'s satisfying (6.1), $\Psi_1(i_2^*)$ is maximum. In general, given $i_1^*, i_2^*, \ldots, i_k^*$ $k = 1,2,\ldots,N-1$ let i_{k+1}^* be such that

$$\frac{\Psi_2(i_{k+1}^*)-\Psi_2(i_k^*)}{\Psi_1(i_{k+1}^*)-\Psi_1(i_k^*)} = \max_{i:\Psi_1(i)>\Psi_1(i_k^*)} \frac{\Psi_2(i)-\Psi_2(i_k^*)}{\Psi_1(i)-\Psi_1(i_k^*)} \tag{6.2}$$

and among all i_{k+1}^*'s satisfying (6.2), $\Psi_1(i_{k+1}^*)$ is maximum. Note that this iteration will end when the point $(\Psi_1(i_N^*), \Psi_2(i_N^*))$ is reached.

Similarly, to graph the lower boundary of $\mathcal{R}$, let

$$\Psi_2(i_1') = \min_{i_1:(\Psi_1(i_1),\Psi_2(i_1))\in\mathcal{R}_1} \Psi_2(i_1) \quad \text{and} \quad \Psi_2(i_N') = \min_{i_N:(\Psi_1(i_N),\Psi_2(i_N))\in\mathcal{R}_2} \Psi_2(i_N)$$

Given $i_1', i_2', \ldots, i_\ell'$ $\ell = 1,2,\ldots,N-1$, let $i_{\ell+1}'$ be such that

$$\frac{\Psi_2(i'_{\ell+1})-\Psi_2(i'_\ell)}{\Psi_1(i'_{\ell+1})-\Psi_1(i'_\ell)} = \min_{i:\Psi_1(i)>\Psi_1(i'_\ell)} \frac{\Psi_2(i)-\Psi_2(i'_\ell)}{\Psi_1(i)-\Psi_1(i'_\ell)} \tag{6.3}$$

and among all $i'_{\ell+1}$'s satisfying (6.3), $\Psi_1(i'_{\ell+1})$ is maximum. This iteration will end when the point $(\Psi_1(i'_N), \Psi_2(i'_N))$ is reached.

Special Case:

Let $\Psi_1(i) = i$ and $\tilde{\Psi}_2(\cdot)$ be the function that results from connecting the points $(i, \Psi_2(i))$ and $(i+1, \Psi_2(i+1))$ for $i = 1,\ldots,N-1$. Then, under some conditions on $\tilde{\Psi}_2(\cdot)$ we find that the probability measure that assigns all its mass to the point $i = n$ solves all the above optimization problems and the optimum value is $\Psi_2(n)$. These conditions on $\tilde{\Psi}_2(\cdot)$ differ from one problem to another. In particular, for problem (1), $\tilde{\Psi}_2(\cdot)$ has to be decreasing and convex while for problem (2), it has to be increasing and convex. For problem (3), it has to be increasing concave and finally for problem (4), it has to be decreasing and concave.

Example 6.1:

In Section 3.1.2, we have faced a special case of problem (1) where $\Psi_1(i) = i$ and $\Psi_2(\cdot)$ is given by

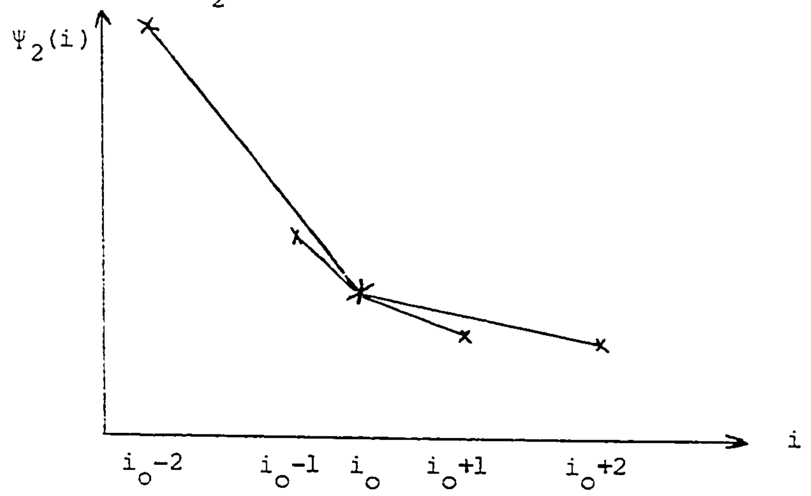
$$\Psi_2(i) = \frac{1}{i}\left(\sum_{j=i+1}^{N} m_{(j)}\right)^2 + \sum_{j=i+1}^{N} m_{(j)}^2$$

where $m_{(1)} \geq m_{(2)} \geq \cdots \geq m_{(N)} > 0$. Now, we will show that $\tilde{\Psi}_2(\cdot)$ is a decreasing convex function on $(0,N)$ and, hence, according to the above special cases, this problem is solved by taking p such that $p(n) = 1$.

It is obvious that $\tilde{\Psi}_2(\cdot)$ is decreasing since

$$\tilde{\Psi}_2(i+1)-\tilde{\Psi}_2(i) = [\frac{1}{i+1}(\sum_{j=i+2}^{N} m_{(j)})^2 + \sum_{j=i+2}^{N} m^2_{(j)}] - [\frac{1}{i}(\sum_{j=i+1}^{N} m_{(j)})^2 + \sum_{j=i+1}^{N} m^2_{(j)}] < 0 .$$

Now, we show that $\tilde{\Psi}_2(\cdot)$ is convex on $(0,N)$



For any arbitrary i_o, where $2 < i_o < N-2$, let L_k denote the slope of the line connecting the two points $(i_o, \tilde{\Psi}_2(i_o))$ and $(i_o+k, \tilde{\Psi}_2(i_o+k))$. We now show that $L_2 \geq L_1$ and $L_{-1} \geq L_{-2}$

$$L_2-L_1 = \frac{\tilde{\Psi}_2(i_o+2)-\tilde{\Psi}_2(i_o)}{2} - \frac{\tilde{\Psi}_2(i_o+1)-\tilde{\Psi}_2(i_o)}{1}$$

$$= \frac{1}{2}\left[\frac{1}{(i_o+2)}\left(\sum_{j=i_o+3}^{N} m_{(j)}\right)^2 + \sum_{j=i_o+3}^{N} m_{(j)}^2 - \frac{1}{i_o}\left(\sum_{j=i_o+1}^{N} m_{(j)}\right)^2 - \sum_{j=i_o+1}^{N} m_{(j)}^2\right]$$

$$-\left[\frac{1}{(i_o+1)}\left(\sum_{j=i_o+2}^{N} m_{(j)}\right)^2 + \sum_{j=i_o+2}^{N} m_{(j)}^2 - \frac{1}{i_o}\left(\sum_{j=i_o+1}^{N} m_{(j)}\right)^2 - \sum_{j=i_o+1}^{N} m_{(j)}^2\right]$$

$$= \frac{1}{2}\left[\frac{1}{(i_o+2)}\left(\sum_{j=i_o+3}^{N} m_{(j)}\right)^2 - \frac{1}{i_o}\left(\sum_{i_o+1}^{N} m_{(j)}\right)^2 - m_{(i_o+1)}^2 - m_{(i_o+2)}^2\right]$$

$$- \left[\frac{1}{(i_o+1)}\left(\sum_{j=i_o+2}^{N} m_{(j)}\right)^2 - \frac{1}{i_o}\left(\sum_{j=i_o+1}^{N} m_{(j)}\right)^2 - m_{(i_o+1)}^2\right]$$

$$= \frac{1}{2(i_o+2)}\left(\sum_{j=i_o+3}^{N} m_{(j)}\right)^2 + \frac{1}{2i_o}\left(\sum_{j=i_o+1}^{N} m_{(j)}\right)^2 + \frac{1}{2} m_{(i_o+1)}^2 - \frac{1}{2} m_{(i_o+2)}^2$$

$$- \frac{1}{(i_o+1)}\left(\sum_{j=i_o+2}^{N} m_{(j)}\right)^2 .$$

Letting $a_1 = \sum_{j=i_o+3}^{N} m_{(j)}$ we get

$$L_2 - L_1 = \frac{1}{2(i_o+2)} a_1^2 + \frac{1}{2i_o}(a_1 + m_{(i_o+1)} + m_{(i_o+2)})^2 + \frac{1}{2} m_{(i_o+1)}^2 - \frac{1}{2} m_{(i_o+2)}^2$$

$$- \frac{1}{(i_o+1)}(a_1 + m_{(i_o+2)})^2$$

$$= [\frac{1}{2(i_o+2)} + \frac{1}{2i_o} - \frac{1}{(i_o+1)}]a_1^2 + [\frac{1}{i_o} m_{(i_o+1)} + \frac{1}{i_o} m_{(i_o+2)}$$

$$- \frac{2}{(i_o+1)} m_{(i_o+2)}]a_1 + [\frac{1}{2i_o} m^2_{(i_o+1)} + \frac{1}{2i_o} m^2_{(i_o+2)}$$

$$+ \frac{1}{i_o} m_{(i_o+1)} m_{(i_o+2)} + \frac{1}{2} m^2_{(i_o+1)} - \frac{1}{2} m^2_{(i_o+2)} - \frac{1}{(i_o+1)} m^2_{(i_o+2)}]$$

$$= [\frac{1}{i_o(i_o+1)(i_o+2)}]a_1^2 + [\frac{1}{i_o} m_{(i_o+1)} - \frac{1}{(i_o+1)} m_{(i_o+2)}$$

$$+ (\frac{1}{i_o} - \frac{1}{(i_o+1)}) m_{(i_o+2)}]a_1 + [(\frac{1}{2i_o} m^2_{(i_o+1)} + \frac{1}{2i_o} m^2_{(i_o+2)})$$

$$+ (\frac{1}{i_o} m_{(i_o+1)} m_{(i_o+2)} - \frac{1}{(i_o+1)} m^2_{(i_o+2)})$$

$$+ (\frac{1}{2} m^2_{(i_o+1)} - \frac{1}{2} m^2_{(i_o+2)})]$$

$$> 0.$$

Similarly,

$$L_{-1}-L_{-2} = [\Psi_2(i_o)-\Psi_2(i_o-1)] - \frac{1}{2}[\Psi_2(i_o)-\Psi_2(i_o-2)]$$

$$= [\frac{1}{i_o}(\sum_{j=i_o+1}^{N} m_{(j)})^2 + \sum_{j=i_o+1}^{N} m^2_{(j)} - \frac{1}{(i_o-1)}(\sum_{j=i_o}^{N} m_{(j)})^2 - \sum_{j=i_o}^{N} m^2_{(j)}]$$

$$- \frac{1}{2}[\frac{1}{i_o}(\sum_{j=i_o+1}^{N} m_{(j)})^2 + \sum_{j=i_o+1}^{N} m^2_{(j)} - \frac{1}{(i_o-2)}(\sum_{j=i_o-1}^{N} m_{(j)})^2$$

$$- \sum_{j=i_o-1}^{N} m^2_{(j)}]$$

$$= [\frac{1}{i_o}(\sum_{j=i_o+1}^{N} m_{(j)})^2 - \frac{1}{(i_o-1)}(\sum_{j=i_o}^{N} m_{(j)})^2 - m^2_{(i_o)}]$$

$$- \frac{1}{2}[\frac{1}{i_o}(\sum_{j=i_o+1}^{N} m_{(j)})^2 - \frac{1}{(i_o-2)}(\sum_{j=i_o-1}^{N} m_{(j)})^2 - m^2_{(i_o-1)} - m^2_{(i_o)}]$$

$$= \frac{1}{2i_o}(\sum_{j=i_o+1}^{N} m_{(j)})^2 - \frac{1}{(i_o-1)}(\sum_{j=i_o}^{N} m_{(j)})^2 + \frac{1}{2(i_o-2)}(\sum_{j=i_o-1}^{N} m_{(j)})^2$$

$$- \frac{1}{2} m^2_{(i_o)} + \frac{1}{2} m^2_{(i_o-1)} .$$

Letting $a_2 = \sum_{j=i_o+1}^{N} m_{(j)}$ we get

$$L_{-1}-L_{-2} = \frac{1}{2i_o} a_2^2 - \frac{1}{(i_o-1)}(a_2+m_{(i_o)})^2 + \frac{1}{2(i_o-2)}(a_2+m_{(i_o)}+m_{(i_o-1)})^2$$

$$- \frac{1}{2} m^2_{(i_o)} + \frac{1}{2} m^2_{(i_o-1)}$$

$$= [\frac{1}{2i_o} - \frac{1}{(i_o-1)} + \frac{1}{2(i_o-2)}]a_2^2 + [-\frac{2}{(i_o-1)} m_{(i_o)} + \frac{1}{(i_o-2)} m_{(i_o)}$$

$$+ \frac{1}{(i_o-2)} m_{(i_o-1)}]a_2 + [-\frac{1}{(i_o-1)} m^2_{(i_o)} + \frac{1}{2(i_o-2)}(m^2_{(i_o)}+m^2_{(i_o-1)}$$

$$+ 2m_{(i_o)}m_{(i_o-1)}) - \frac{1}{2} m^2_{(i_o)} + \frac{1}{2} m^2_{(i_o-1)}]$$

$$= [\frac{1}{i_o(i_o-1)(i_o-2)}]a_2^2 + [(\frac{1}{(i_o-2)} - \frac{1}{(i_o-1)})m_{(i_o)}$$

$$+ (\frac{1}{(i_o-2)} m_{(i_o-1)} - \frac{1}{(i_o-1)} m_{(i_o)})]a_2$$

$$+ [(-\frac{1}{(i_o-1)} m^2_{(i_o)} + \frac{1}{(i_o-2)} m_{(i_o)}m_{(i_o-1)})$$

$$+ \frac{1}{2(i_o-2)}(m^2_{(i_o)}+m^2_{(i_o-1)}) + \frac{1}{2}(m^2_{(i_o-1)}-m^2_{(i_o)})]$$

$$> 0 .$$

<u>Example 6.2</u>:

In Section 3.1.2, we have met another special case of problem (1) where $\Psi_1(i) = i$ and $\Psi_2(\cdot)$ is given by

$$\Psi_2(i) = \sum_{j=i+1}^{N} m^2_{(j)} \quad i = 1,\ldots,N \quad \text{and} \quad \Psi_2(N) \equiv 0$$

where $m_{(1)} \geq m_{(2)} \geq \cdots \geq m_{(N)} > 0$. We now show that $\tilde{\Psi}_2(\cdot)$ is a *decreasing convex function on* $(0,N)$ and, hence, according to the above special cases, this problem is solved by taking p such that $p(n) = 1$.

It is obvious that $\tilde{\Psi}_2(\cdot)$ is decreasing since

$$\tilde{\Psi}_2(i+1) - \tilde{\Psi}_2(i) = \sum_{j=i+2}^{N} m^2_{(j)} - \sum_{j=i+1}^{N} m^2_{(j)} < 0.$$

As in Example 6.1, for any arbitrary i_o where $2 < i_o < N-2$, let L_k denote the slope of the line connecting the two points $(i_o, \tilde{\Psi}_2(i_o))$ and $(i_o+k, \tilde{\Psi}_2(i_o+k))$. To show that $\tilde{\Psi}_2(\cdot)$ is convex on $(0,N)$, it is enough to show that $L_2 \geq L_1$ and $L_{-1} \geq L_{-2}$ as follows:

$$L_2-L_1 = [\frac{\tilde{\Psi}_2(i_o+2)-\tilde{\Psi}_2(i_o)}{2}] - [\frac{\tilde{\Psi}_2(i_o+1)-\tilde{\Psi}_2(i_o)}{1}]$$

$$= \frac{1}{2}[\sum_{j=i_o+3}^{N} m^2_{(j)} - \sum_{j=i_o+1}^{N} m^2_{(j)}] - [\sum_{j=i_o+2}^{N} m^2_{(j)} - \sum_{j=i_o+1}^{N} m^2_{(j)}]$$

$$= \frac{1}{2} \sum_{j=i_o+3}^{N} m^2_{(j)} + \frac{1}{2} \sum_{j=i_o+1}^{N} m^2_{(j)} - \sum_{j=i_o+2}^{N} m^2_{(j)}$$

$$= - \frac{1}{2} m^2_{(i_o+2)} + \frac{1}{2} m^2_{(i_o+1)} \geq 0.$$

Similarly,

$$L_{-1}-L_{-2} = [\Psi_2(i_o)-\Psi_2(i_o-1)] - \frac{1}{2}[\Psi_2(i_o)-\Psi_2(i_o-2)]$$

$$= \left[\sum_{j=i_o+1}^{N} m^2_{(j)} - \sum_{j=i_o}^{N} m^2_{(j)}\right] - \frac{1}{2}\left[\sum_{j=i_o+1}^{N} m^2_{(j)} - \sum_{j=i_o-1}^{N} m^2_{(j)}\right]$$

$$= \frac{1}{2} \sum_{j=i_o+1}^{N} m^2_{(j)} - \sum_{j=i_o}^{N} m^2_{(j)} + \frac{1}{2} \sum_{j=i_o-1}^{N} m^2_{(j)}$$

$$= -\frac{1}{2} m^2_{(i_o)} + \frac{1}{2} m^2_{(i_o-1)} \geq 0.$$